

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 28, 2005, 20:55:08 ; Search time 52 Seconds
(without alignments)
2697.415 Million cell updates/sec

Title: US-10-607-631-8

Perfect score: 9732

Sequence: 1 MNKKSTLLATAAAIIGST.....TNNAFNNVKEFNISKIVE 1879

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 513545 seqs, 74649064 residues

Total number of hits satisfying chosen parameters: 513545

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents AA.*

- 1: /cgn2_6/ptodata/1/iaa/5A-COMB.pep.*
- 2: /cgn2_6/ptodata/1/iaa/5B-COMB.pep.*
- 3: /cgn2_6/ptodata/1/iaa/6A-COMB.pep.*
- 4: /cgn2_6/ptodata/1/iaa/6B-COMB.pep.*
- 5: /cgn2_6/ptodata/1/iaa/PCTUS-COMB.pep.*
- 6: /cgn2_6/ptodata/1/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	301.5	3.1	2662	4	US-09-595-684B-31
2	301	3.1	2663	4	Sequence 31, Appli
3	286.5	2.9	1979	4	Sequence 1252, Ap
4	284.5	2.9	2047	4	Sequence 6468, Ap
5	276	2.8	3259	4	Sequence 7404, Ap
6	275.5	2.8	1663	5	Sequence 6507, Ap
7	275.5	2.8	10182	3	PCT-US93-07261-16
8	271	2.8	1733	3	Sequence 1159, Ap
9	262.5	2.7	5024	4	Sequence 11433, A
10	256.5	2.6	1588	5	Sequence 2964, Ap
11	256.5	2.6	3696	3	Sequence 11, Appl
12	255	2.6	3418	3	Sequence 5080, Ap
13	255	2.6	2482	3	Sequence 4, Appl
14	253	2.6	3418	2	Sequence 6, Appli
15	253	2.6	3418	3	Sequence 2, Appli
16	253	2.6	3418	3	Sequence 2, Appli
17	252	2.6	2285	3	Sequence 2, Appli
18	252	2.6	2285	4	Sequence 2, Appli
19	250.5	2.6	2710	1	Sequence 6, Appli
20	250.5	2.6	2710	1	Sequence 6, Appli
21	250.5	2.6	2710	3	Sequence 6, Appli
22	250.5	2.6	2710	3	Sequence 6, Appli
23	250.5	2.6	2710	4	Sequence 6, Appli
24	250.5	2.6	2710	4	Sequence 6, Appli
25	250.5	2.6	3418	2	Sequence 4, Appli
26	250.5	2.6	3418	3	Sequence 4, Appli
27	250.5	2.6	3418	3	Sequence 4, Appli

28	245	2.5	904	3	US-09-198-484-2	Sequence 2, Appli
29	244.5	2.5	2329	3	US-08-755-587-16	Sequence 16, Appli
30	242.5	2.5	2954	4	US-09-150-867-1	Sequence 1, Appli
31	238.5	2.5	3913	4	US-09-949-016-10933	Sequence 10933, A
32	238.5	2.5	4377	4	US-09-949-016-6978	Sequence 6978, Ap
33	233	2.4	2807	4	US-09-543-681A-4980	Sequence 4980, Ap
34	231	2.4	3210	4	US-09-538-092-1154	Sequence 1154, Ap
35	231	2.4	3878	4	US-09-914-269-11	Sequence 11, Appl
36	229.5	2.4	2472	4	US-09-538-092-1312	Sequence 1312, Ap
37	228	2.3	3248	1	US-08-353-700-1	Sequence 1, Appli
38	228	2.3	3248	5	PCT-US95-16216-1	Sequence 1, Appli
39	227.5	2.3	1010	3	US-09-134-001C-5178	Sequence 5178, Ap
40	227	2.3	2777	4	US-09-543-681A-6124	Sequence 6124, Ap
41	226.5	2.3	3135	1	US-08-323-170B-2	Sequence 2, Appli
42	226.5	2.3	3135	3	US-08-954-441-2	Sequence 2, Appli
43	226	2.3	1935	4	US-09-538-092-916	Sequence 916, App
44	225.5	2.3	2186	4	US-09-949-016-10828	Sequence 10828, A
45	225	2.3	1944	4	US-09-949-016-10929	Sequence 10929, A

ALIGNMENTS

RESULT 1

US-09-595-684B-31

; Sequence 31, Application US/09595684B

; Patent No. 654766

; GENERAL INFORMATION:

; APPLICANT: Beraud, Christophe

; APPLICANT: Ohashi, Cara

; APPLICANT: Sakowicz, Roman

; APPLICANT: Vaisberg, Eugeni

; APPLICANT: Wood, Kenneth

; APPLICANT: Yu, Ming

; TITLE OF INVENTION: Human kinesins and methods of producing

; FILE OF INVENTION: and purifying human kinesins

; FILE REFERENCE: cytop036

; CURRENT APPLICATION NUMBER: US/09/595,684B

; PRIOR FILING DATE: 2002-06-24

; PRIOR APPLICATION NUMBER: 09/295,612

; PRIOR FILING DATE: 2000-04-20

; NUMBER OF SEQ ID NOS: 105

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 31

; LENGTH: 2662

; TYPE: PRT

; ORGANISM: Human

US-09-595-684B-31

Query Match 3.1%; Score 301.5; DB 4; Length 2662;
Best Local Similarity 19.1%; Pred. No. 2.7e-11;
Matches 366; Conservative 330; Mismatches 676; Indels 549; Gaps 92;

Qy	1	MNKKSTLLATAAAIIGSTVGTGVLGSLKVKYGVNPTQGVISQGLIDSVAFKPSIA	60
Db	763	IQDKSEELHIITSEK---DKLFSEVHKSEKSV-----QGLLEEIG-----RTKODLA	806
Qy	61	NFTSDYQSVKALLNGKTFDPKSEFTDFVSKPFLTNGRTVLEIPKQYQVISEFSPE	120
Db	807	TTQSNYKSTDOEQFNKTLH-----MDFEQKYQVLEE-----	839
Qy	121	DDKERFRGLPHLKEKLEDGNTAQSAKTFIVLL-----PLDMPKAALGQSYIVDKNFNL	175
Db	840	---NRM-----NQELVNLSEKQAQFDSLGALKTSLSYKTQELQKTEVQERLNM	889
Qy	176	--IITHPLSNFSAQSIKPL-ALTRSSDFIA-KLNFFNNODELMVLYLEKFFDLEALKANIRL	231
Db	890	EQLKEQLENRDS---PLQTVREKTLITEKLOQ--TLEEVTKLTQEKDDLKQLQSLQI	943
Qy	232	QTADFSFEKGNLVPFVYSFIRNPQNKQEWASDLNQDKTVRLYLRTFEFSQAATILKDY	291
Db	944	E-----RDQLKSDIHD-----TVNMNIDTQ--EQLRNALES	973

Db 763 IQDKSELHIITSEK---DKLFSEVHVHESRV-----QGLLEEIG-----KTODDLA 806
Qy 61 NPTSDYQSVKALLANGTKFPDKSEFTDFVSKFDELNNORTVLETPKTKYQVVISFSP 120
Db 807 TTQSNYKSTDOEQFNPKTLH-----MDPEQKYKMYLEE----- 839
Qy 121 DDKERPLGFLHKEKLEDGNIAGSATKFIYLL-----PLDMPKAAALGOYSYIVDKNPNL 175
Db 840 --NEM-----NQIVNLSKBAQFDSGLGALTKELSYKQEQLOEKTRFQOERLNE 889
Qy 176 --IHLPLNSAQSICKPL-ALTRSSDEIA-KLAFNNQDELWYVLEKFFDLEAKANIRL 231
Db 890 BQLEKENRDS---PLQVREKTLITKELQ--TLEEVKTLTQEKDDLKQLOESLQI 943
Qy 232 QTADFSPEKGNLDPFVYSFIRNPQNKWASDLNQDKTVRLYLRTEFSPQATIL--- 288
Db 944 ERDQL--KSDIHT-VNMNIDTQOELRNALLESKQHQETINT-LKSKISEEVSRLHME 998
Qy 289 -KDYKYDETFLLSI-----DLKASGTSLFAN-----ENDL--- 319
Db 999 ENTGETKDEFQOKWVGIDKKODLEAKNTQTUTADVKNONEIEBQOKIFSLIOEKNELOQM 1058
Qy 320 -----KOOLDVLLD-----VSDYFGGQSETIT--SNSQVKKVPASERS 356
Db 1059 LESVIAKEQLKTDLENIENTIENQBELRLGLDELKKQOEIVAQEKNAHAIKKEGELSRT 1118
Qy 357 LKORVFKQDQOKPRIEKSFYDYDALSFYSQLOELVSKPNSIKDLVNATLARNLRSIG 416
Db 1119 C-DRLAEVEEKLEKSOQLOEQOQLLNVQEMSEMOKKINEIENLKNE--LKNKELTLE 1175
Qy 417 KYNFLPDDLASHLDYPLVLSKAKIKQSSITKFLIEUFIKISLASSILGQEPNIKTLFE 476
Db 1176 HMETERLELAQKLENBVEVKSITKERKVLKEL-----QKSFE 1213
Qy 477 KEVTFKLDNPRDVEIEKAFGLYLPVGBEELQARKAPASPEKSKKGLKEFQOKEEN 536
Db 1214 TERDHLRGYIREIE--ATGL--QTKELALIAH-----IHLKEHQETIDELARSVEK 1261
Qy 537 SKAINNOGBLEEDNITERLPENSPIOYQOENAGLGASPDPMKQVONQRYLAKSQI 596
Db 1262 TQAINTQDLEKSHST--KLQEEIPVLHEEQEL-----LPNVKQVSETOETWNELELL 1311
Qy 597 QELIKADYTKLAKLSNRHNTYISLRLEKQOLFQVNPRIPS--SRDIEKAKEVLDKTEKN 655
Db 1312 TEQSTTKDSTTLTIEMER-----LRLNEKQESQEEIKSLTERDMNLTKIKALEV-K 1364
Qy 656 YWQIYSSASPVFQKWSLFGYVRYLLGLDPKQTIHE-LVKLGQKAGLQFEGYENLPSPFN 714
Db 1365 HDQL-----KEHIRETLAKIOESQSKQ-----EQSLN 1391
Qy 715 LEDLKNIRIKTPLFSQKDNFKLSLDFNYYDGEIKAPFGLPLPLPKELRNSNSGGS 774
Db 1392 MKEKDN--ETTKIVSEMEQFKP-----DSALLRIEIEH-LGLSKRLQESH-- 1435
Qy 775 QNSNSPWEQEIIISQDKQNLNODOLAQFSTKIWE---KIIGDENEFPQNNRLOYKLKLD 831
Db 1436 -----EMKSVAKESKDLQRLQEQVQSESQDLQENIKIEIVANHLETEELKVAHCCLKE 1488
Qy 832 LQESWINKTRDNL-----YWTYLGDKLAKPKPNNLEAK-----FRQISNLQBL 875
Db 1489 -QETINELRVNLSKEYTEISTIQOLEAINDKLQNKIOEIVKEEQLNKOISEVQENV 1547
Qy 876 TAF-----YTSALSNMNYQDSGAKSTIIEETAEALDPKPKVEKGVADV 920
Db 1548 NELQFKHEKRAKADSALQIESKMLELTNRLQESQEEIQIMTEKEEM-KRVOEALQIER 1606
Qy 921 YOLKHYAIGFDNAGKNQFQVIRSSRTI-YLKTSGKSLKLEAD--TIDQLNQAVKNAPL 977
Db 1607 DQLK-----ENTKEIVAKKESQEKYQFLKMTAVNETQEKWCIEHLKEQETOKL 1658
Qy 978 GLQSFYLDTERFGVFOKATSLAVQH-----KQEKTLPKKLNNNDGYTLIHLKLLK 1028
Db 1659 NLEN--IETENRILQILHENLEEMRSVTKERDRLRSVEETL--KVERD---QLKENLRE 1711

Qy 1029 PVTQISSSPKDWFEKGLNQSQSONVNVSTFGSIIESPSTNPFQEDADLDQD--- 1084
Db 1712 TTRDLEKQEBELKIVMHLKEH---QETIDKLGIVSEKTNESNMQKLEHNSDALKAQ 1768
Qy 1085 ----QDSSRQGNNSLDNQE-----AGLLKQKLAILL-----GNQFIYYQONDKEI 1126
Db 1769 DLKIQEBELRAHMLKEQOETIDKLGIVSEKTDKLSNMQKOLENSNAKLOEIQELKAN 1828
Qy 1127 EPEIINV-----EKVSELSFRVFKLAKTLDNGKTRIVLSDETMSLIVNTTIEKTP 1178
Db 1829 EHOLITLKXDVNSTQKVSME-----OLKQKIQOSLTSKLEIENLNI-AQELHENLE 1882
Qy 1179 EMSAVEPEVPT-KWVQ-----YDPRTPLAAKTKFVLKFKDQIPVDG 1219
Db 1883 EMKSVMKERDNRVRBETLKLERDQLKESIQETKARDLEIQOELKTARMLSKHEKETVD- 1941
Qy 1220 SGNISDKWLASIPLVTHQOMLRSLPVVVTIRELGLTKTEQOQOQOQOQOQOQKAVRKE 1279
Db 1942 --KLREK-----ISEKTIQISDIQK-----DLQSKDELOKQIQOELQKELQLLRVK 1986
Qy 1280 EELETYNPKDEFNLPKTKAHLTLNLVNDPNYKIEDLKVIKNEAGDHLAFSLRAN 1339
Db 1987 EDVNMSSHK--INEMBOLKKQF-----EPNY-----LCKCENDNFQLTKKLH-- 2026
Qy 1340 NIKRLMNTPTITFADYNPFFYINEDWRSIDKYLNNKGN--VSSHQQAAGNQSSGLIQL 1397
Db 2027 --ESLBEIRIVAKE-----RDELARIKESLAWERQFIATLREMIARDRQ----- 2069
Qy 1398 NKNIKPTFTFPALIAKDRNTNMLSNYSKDIIMIKPKYLVERSIGVPSWTGLDGYIGSEQ 1457
Db 2070 NHQVKPEK-----RLISDGOQHLMESLREKCSRIKE--LLKR-----YSEMDHY----- 2112
Qy 1458 TKDGTSSSSQOKGFDQFIOALGI---KNTVEYHG-----KLGLSIRIFDPGNEAKIKDA 1509
Db 2113 -----ECLNRLSLDLEKEIEFHRIMKKLKVLSY-----VTKLEE 2148
Qy 1510 SNKKGBEKLKSYDLFKNYLNEVEKSPKIAKAGWTNIHPDQKEYPNPNQKLPENYLNVL 1569
Db 2149 -----QHECINKPEM--DFIDEVEKQKELLIK---IQHLQOCDQDVPISRELRL--LKL 2193
Qy 1570 NQPMKVTL-----YNSSDP-----ITN-----LFVPEGSGRSGTKLKQVIOK 1608
Db 2194 NQNDLHIEELKDFSESEFFSIKTEFQVLSNRKEMTQFLEBWLNTRFDIEKLKNGIQK 2253
Qy 1609 -----QVNNYADWGSAYLTFWYDKNI--ITNQPNVI---TANIADVFIDVKEL-EDNT 1657
Db 2254 ENDRICQVNN-----FFNRIIAIWNESFEEERSATISKWEQDLKSLKEKNE 2302
Qy 1658 KLIAPNITQWPNPISGSKPKYKPTVPFGNWNENSSNMNSQAQTPPTWEKIRE 1709
Db 2303 KLFKNYQTLKTSLASGAQ---VNPFT---TQDNKNPHVTSRATQLTTEKIRE 2347

RESULT 3

US-09-949-016-6468
; Sequence 6468, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0

```
; SEQ ID NO 6468
; LENGTH: 1979
; TYPE: PRT
; ORGANISM: Human
US-09-949-016-6468

Query Match      2.9%; Score 286.5; DB 4; Length 1979;
Best Local Similarity 19.0%; Pred. No. 1.9e-10;
Matches 373; Conservative 345; Mismatches 682; Indels 563; Gaps 99;

QY 22 FGVVGLASVYKRVGNPTQGVISQLGLSDSVAFKPSIANFTSD-----YQSVKKAL 73
Db 164 FGDIISSQGEIN-RLSNEVSRLESEVGHWRHTA-QTSKAGQTDNSDQSEICKLQNIKEU 221
QY 74 LNKGTDPKSEPTDFVSPDFLTN-NGRTVLEIPKKYQVVISSEFDEPKFRIGPHL 132
Db 222 KQN-----RSQIEDDHQHEMSVLQNAHQKQLTEISRRHREELSDY-----EERI----- 265
QY 133 KEKLEDCNTAQSATKIYLLPLDMPKAAALGOYSYIVDKNFNNLIHPLNFSQAQSIKPLA 192
Db 266 -BELE--NLQOQSGV--IETDLSKI-----YEMOKTIQVLIQEKVE--STKQMEQLE 312
QY 193 LTRSSDFIAKLNQFN-----QDELWYLEKFPFDEALKANIRLQ-----TA 234
Db 313 -DKIDINKKLSAENDRIILRREQQLNV--EKROIIECE--NLKLECSKLQPSAVKQS 368
QY 235 DFSFEKGNLV-----DPPF-VYSFIRNPQNKEMASDLNQD-----QKTVRILYLRTEFSP 282
Db 369 DTWTEKERILAQASAVEEVRFLQALSDAENEIMRLSSLNQDNSLAEDNLKMKRIEVL 428
QY 283 QAKTILKYKDYKDETFLSSIDLK-----ASNGTSLPANENDLKQDLDVDDLVDSDY 333
Db 429 KEKSLLS--QEKELQMSLLKLNNYEYIKATATRIDISLDELHLRLNLE----- 477
QY 334 FGGOSSETITSQVQVPASERSLKDVRKFKDQKQPRIEKFSLYEYDALSFYSOLOEVL 393
Db 478 -----AKEQELNOSIS-EKETLAEIEELDRQNEQATKMILIKOOL 518
QY 394 SKPNSIKDLVNATLARNLRFSLGKYNFLFD---DLASHLDYDFYFLVSKAKIKOSSITKULF 450
Db 519 SKQNEGDSIISKQLDNLDEKXRVHQLEDKMDITKELD---VQKELIQSEVA--- 570
QY 451 IELPIKISLSKSSILGQDEPNIKTLPEKVTFKLDNFRDVRVEIEKAFGLLYPGVNEELEQAR 510
Db 571 -----LNDLH-----LTKOKLEDKVENL-----VDQLN 593
QY 511 KAORAFSEKSKKGLKEFSQOKEENSKAINNQ--EGLBEDDN-----ITER----- 555
Db 594 KSQESNVSTOKENLELKEHIRQNEBELSRIRNELMQSLNQDNSNFKDTLLKEREAVERN 653
QY 556 LPEN-SPIQYQOBNAGLGASDPKPYMIKDVQNRYYLA----KSQIQELIKAKDYTKLAK 610
Db 654 LKQNLSELEQNLNKKVAPDVK-----MENEKLVLACEDVRHQLEECLAGNQLSLEK 707
QY 611 LLSNRHTYNISLBLEQOLFQVNDPNRIPSSRDIKAKFVLDKTEKNKYQWYISASVPFQNK 670
Db 708 -----NTIVETLKMEGEIEAEALCWAKRLLEEA-----NKYEKTEIELSNA-RN- 751
QY 671 WSLFGYVRYLLGDDPKQTHIE-LVKLGAKAGLQF-----EGYENLRPSDF-----NLE 716
Db 752 -----LNTSALQLEHEHLIKLNQKMDIEALKNIEQOMDTDHKETKDVLSSSLE 801
QY 717 DLKNIRIKTPLFSQDNFKLSLDDFNYYDGEIKAPEFGLPLPLPKELRRN----- 767
Db 802 EQQL---TOLINKKEIFIEKLEKRSKQLEELDK-----YSQALRKNELIRQTIE 850
QY 768 SSSNGSGQNSNSPWEQBIISQFKDQ-----NLNQDOLA-----QFST----- 805
Db 851 KORSGLSMKEENHNLQELERLEEQSRAPVADPKTLDSTVTELAISEVSQLNTIKEHLEE 910
QY 806 --KIWEKIIGDENEFQDNRLQYLLKDLQESWINKRDLNLYWYLGDKLVKPKQNLBA 863
Db 911 EIKGHQKIIEDNQ-----SKWQ--LLQSLQEQ--KKEMDE--FRYQHEQMNATHQTLFLE 960
```

RESULT 4

US-09-949-016-7404

; Sequence 7404, Application US/09949016

Patent No. 6812339		Query Match		2.9% Score 284.5; DB 4; Length 2047;	
GENERAL INFORMATION:		Best Local Similarity		18.9% Pred. No. 2.7e-10;	
APPLICANT: VENTER, J. Craig et al.		Matches 370; Conservative 346; Mismatches 686; Indels 559; Gaps 97;			
TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED					
TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF					
FILE REFERENCE: CL001307					
CURRENT APPLICATION NUMBER: US/09/949,016					
CURRENT FILING DATE: 2000-04-14					
PRIOR APPLICATION NUMBER: 60/241,755					
PRIOR FILING DATE: 2000-10-20					
PRIOR APPLICATION NUMBER: 60/237,768					
PRIOR FILING DATE: 2000-10-03					
PRIOR APPLICATION NUMBER: 60/231,498					
PRIOR FILING DATE: 2000-09-08					
NUMBER OF SEQ ID NOS: 207012					
SOFTWARE: FastSeq for Windows Version 4.0					
SEQ ID NO 7404					
LENGTH: 2047					
TYPE: PRT					
ORGANISM: Human					
US-09-949-016-7404					
QY	22	FGTVVGLASKVKYRGVNPQTGVI	SQLGLIDSVAFKPSIANTPST	-----YQSVKAL	73
DB	232	FGDIISQQEIN-RLSNEVSRLESEVGHWHIA-QTSKAQGTDSQDSEICKLQNI	IKEL	289	
QY	74	LNGKTFDPKSEFTDFVSKFDELTN-NGRTVLEIPKYYVVISFSPEDDKERFRLGHL	132		
DB	290	KQN-----RSQIDHQBHEMSVLQNAHQOKUTEISRRHRELSDY-----BERI	333		
QY	133	KEKLEDGNIAGSATKFIYLLPLDMPEKAALGOYSYIVDKNFNLIHPLSNFSAOSIKPLA	192		
DB	334	-BELE--NLLOQGGGV--IETDLSKI-----YEMQTIQVQIEKVE--STKMEGLE	380		
QY	193	LTRSDFFIAKLNQFN-----QDELWVLEKFFDLBALKANIRLO-----TA	234		
DB	381	-DKINDINKLSSANDRDLRREQEQLNV--EKQIMEECE-NLKLECSKLOPSAVKQS	436		
QY	235	DFSPFKGNLV-----DPF-VYSIRNPQOKEWASDLNOD-----QKTVRLYLRTFSP	282		
DB	437	DTWTEKERILAQASVEEVFRLQQALSDAENIEMRLSLNODNSLAEDNLKLMKRIEVL	496		
QY	283	OAKTILKDYKDETFLLSIDLK-----ASNGTSLFANENDLKQDLVDLDDVSDY	333		
DB	497	KEKSLLS--QEKELQMSLLKLNNEYEVIKSTATRDISLDELHDLRLNLE-----	545		
QY	334	FGQSETITNSQVFPVPAASRLKDRVKFKDQOKPRIEKPSLYEYDALSFYSQLOELV	393		
DB	546	-----AKEQELNQSIS-EKETLIAIEELDRQNEATKHMILIKQOL	586		
QY	394	SPNSIKOLVNTATLARNLRFSLGKYNFLD-----DLASHLDYFLYSKAKIKOSSITKLP	450		
DB	587	SKQONEGDSIISKQLDNDEKRVHQLEDDDKMDITKELD-----VQEKLIQSEVA-----	638		
QY	451	IELPIKISLSSITLGDQEPNIKTLPKEVTFKLDNFRDVEIEKAFGLLYPGVNEELEQAR	510		
DB	639	-----LNDLH-----LTQKLEDKVENL-----VDQLN	661		
QY	511	KAQRASFEKSKKGLKEFSQOKEENSKAINNQ-----EGLEEDN-----ITER	555		
DB	662	KSQESNVSIQKENLELKEHRIQNEEELSRIRNELMQSLNQDSNPNKOTLLKEREAEVRN	721		
QY	556	LPEN-SPIQVQOQBNAGLGHSPDKPYMIKDVQVQRYLA-----XSQIOELIKAKDYTKLAK	610		
DB	722	LKQNLSELEQLNENLKKVAFDVK-----MENEKVLACEDVRHQLEECLAGNQLSLEK	775		
QY	611	LLSNRHTYNIISLRLKEQLFDVNPRIIPSSRDIEKAKFVLDTKTKNKYQWQIYSSASPVFQNK	670		
DB	776	-----NTIVETLQWKEGEIEAELCWAKKRLLEA-----NKYEKTIELSNA-RN-	819		

QY	671	WSLFGYVYLLGLDLPKQTIHE-LVKLGOKAGLOF-----EGYENLPDSF-----NLE	716
DB	820	-----LNTSALOQLEHEHLIKLNQKQMEIAELKKNIEQMDTDHKETDVLSSSLE	869
QY	717	DLKNIRIKTPLFSQKNFKLSLLDFNNYVDGEIKAPFGLPLFLPKELRN-----	767
DB	870	EQQL-----TQINKKEIFIEKLKERSKLQEELDK-----YSQALRKNELRTOTIEE	918
QY	768	SSNSGGSQNSNPWEQEIISQFKDQ-----NLSNQDLA-----QFST-----	805
DB	919	KDRSLGSMKEENHLOEELERLKEEORSRTAPVADPKTLDSTVTELASVSLNTIKEHLEE	978
QY	806	--KIWEKIIDENEFQNNRLQYKLLKDLQESWINTRDNLNLYTYLGDKLKVPKNNLEA	863
DB	979	EIKHQKIIEDQNO-----SKMQ--LLQSLQEQ--KKEMDE--FRYQHEQMNATHQTQLFLE	1028
QY	864	KFRQISNLQELLTAFVT-----SAALSNNWNYQDSGAKSTI	900
DB	1029	KDBEIKSLQKTIQIKTQLHEERQDITQNSDIFQETKVOQLNIENGSEKHDLSKAETER	1088
QY	901	IFEEIAELDPKVK--EK--VCADVYOLKFHYAIGFDNAGKFNQEV-----	942
DB	1089	LVRGIKERELEIKLNEKNISLTQIDQLS-----KDEVGKLTQIIQOKDLEIQALHA	1141
QY	943	-IRSSRT--IYLKTSKSKLEADTIDQ-----LNQAVQNAFLGLQSFYLDTERFGVF	992
DB	1142	RISTSTQDVVYL-----QOOLQAYAMEREKVFVNLNEKTR-----ENSHLKTEYHKWM	1191
QY	993	QKLATSLA--VOHQKQKETLPKKNLNDGYTLIHDKLLKPVIPQISSPPEKOWFEGKLNQN	1050
DB	1192	DIVAAKEAALIKLDENKSLSTRFESSGQDMFRETIGN--LSRIIR--EKDIEIDALSQK	1247
QY	1051	QOS-----QNVNVTSGSIIESPYFSINFQEDADLD-----QDGDODS	1088
DB	1248	COTLLAVLOTSTGNEAGVNSNQFEELQERDKLQOVKMEEWKQOVMTTVQNMQES	1307
QY	1089	ROGNNSLDNQBAGLL-----KQKLAILLGNQFIQYQOQNDKEI-----	1126
DB	1308	AQLOEHLQLOAQVLDVSDNNSKLOVDY--TGLIQSYEQNETKLNKQFQELAQVQHSIGQL	1366
QY	1127	-----BFEIIN-----VEKVSSELSFRVEFKLAKTLED	1153
DB	1367	CNTKDLGLGLDIIISPOLSSASLLTPQSAECLRASKSEVLSSESSELQOELLEELRSIQE	1426
QY	1154	NGKTIRVLSDE-----TWSLIVNTTIETKTPMSAVPEVFTKMWQYDPTPLAAKTFVL	1209
DB	1427	KDATIRTLOENNHRLSDSIAATSELERKEHEQTDSIEIKQLK--EKQDVQLKLLKEKULLI	1484
QY	1210	KFKDQIPVDGSGNISDKW-----LASIPLVIHQOQMLRL-----SPVVKTIRELG	1253
DB	1485	KAKSDQLSSNENFTKNVENELLRQAVTMLKERILILEMDIGLKGENEKIVETYR--G	1542
QY	1254	LKTEQOQOQOQOQOQOQOQKQKAVKBELEFYNPKEBFINLPLTKAHLRTLNLVNDP	1313
DB	1543	KETEVALQETNMKFSM-----MLREKEFECHSMKKEKALAPEQLLKEKEQKTGELNQLL	1597
QY	1314	N--VKIEDLKVINEAGDHQALFSLRANNIKRLMNTPTITPADYNPPFFYFNEDMSIDKVL	1371
DB	1598	NAVKSQOEKTVVFOQERD-QVMLALKQ--KOMENTAL-----QNEVQRURDK--	1641
QY	1372	NNKGNVSHQOQAAGGCGSLIORLNKNI--KPEITPTALIALKDRNNTNLSNYSDKII	1429
DB	1642	-----EFRSNQE-----LERLNHLLESSEDSYTREALAAEDRE----AKLRKVT	1682
QY	1430	MIKPYLVERSIGVPWSTGLDGYIGSQTKDGTSSSQKQGFDDFIO-----A	1478
DB	1683	VLEEK-LVSSSSNAME-NASHQASVQVESLQOLANVWSKQR--DETALQLSVSQEQVKQYA	1738
QY	1479	LGLKQTEYHGKLGSLIRIFDP-----GNEIAKIKD--ASNKKEEKLKSYDLFPKNYLN	1530
DB	1739	LSLAN-----LQWLEHFOOEKAWYSAELEKQQLIAEWKQNAENLEGKVISLQECID	1792

QY 1355 NPFYNNEDWRSIDKYLNNKG--NVASHQOQAAG-----GNQSGSLQRLNKNKTKPET 1405
Db 1450 -----NTASKGQONTGLKNAPNERQONTGLKNTPEGQONTGL-----KN----- 1489
QY 1406 FTALIALKDRNNTNLSNYSYDK-----IIMIKPYLVERSIGVPWSTGLDGYIGSBOTKD 1460
Db 1490 -TPS-----EGQONTGLKNAANKQONTGLKNTPE-----SBGP-NTGLKNTPEGQONT 1537
QY 1461 G-----TSSSSQ-----KGFDDFTQALGLKNTYHGLGLSIRIFDPGNEELAKIKDASNK- 1512
Db 1538 GLKNTPEGQONTGLKNTPEGQONTGLKNTPEGQONTGLK-NTPEGQONTDLKNA 1595
QY 1513 -KGEELKLSYDLFPKYNLEYEKKSP-----KIAKGTNIHPDQ---KEYPNPKQLP 1561
Db 1596 SKGQONNDLGLKHPNQOGKHTELNNKLNKPKPTDGLKNVKKDELSDNESSDNEKSK 1655
QY 1562 EN 1563
Db 1656 KN 1657

RESULT 7
US-09-134-001C-3159
; Sequence 3159, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134, 001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064, 964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055, 779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 3159
; LENGTH: 10182
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-3159

Query Match 2.8%; Score 275.5; DB 3; Length 10182;
Best Local Similarity 18.5%; Pred. No. 1.2e-08;
Matches 375; Conservative 319; Mismatches 758; Indels 573; Gaps 93;
QY 74 LANGKT-----PDPKSSEPTDFVSKEDFLTNNGRTVLEIPKKYQVVISSEFSPEDDKERF 127
Db 3189 LMGMTQESINNYTKRREAQNTASSADTTINNGDASIE-----QITEN----- 3231
QY 128 LGPHLEKLEDGNIAQSATKFIYLLPLDMPKAALGOYSYIVDKN-----FNNLIH 178
Db 3232 -----KIRVEATNALNEAKHLTAUTSLKTEVRKLSRRGDTNNKKPSSVAYNN-TIH 3285
QY 179 PLSNFAQSITKPIALTRSDFIKLNQFNQNBELWVYLEKFFDLEALKANIRLQTADF 238
Db 3286 SLOSEITQT-ENRANTIINKPIRSVEEVNNAH-----EVNQLNRLT----- 3327
QY 239 EKNLVDPFVYSFIRPNQKQWASDL-NQDQKTVLRLTEPSPQAKTILKDYKYKDET 297
Db 3328 DTINLLQK-----LANKESLUKEARNLESKINETVQTDGTOQS-----VENYK----- 3371
QY 298 FLSSIDLKASNGTSL---FANENDLKDQ-----LDVDLLDVSDFGQSETTITSNQ 346
Db 3372 ---QAKIAQNESSIAQTLLINGDASDQSVSTEIEKLNKLSLTN-----SINHITVKE 3424
QY 347 VKPVPASERSLKDRVKFKDQKPKRLEKPSFLVYDALSPYQQLQELVSPKSPNSIKOLVAT 406
Db 3425 --PLETAKNQLQANI-----DQKPSD--GMTQSQSVQSERKIQEAKDKINSI-----NNV 3471
QY 407 LARNLFLSLGKYNFLFDDLASHLDYFLVSKAKIKOSSITKKLFIELPIKISKSILGD 466

Db 3472 LANNPDVNAIRTN-----KVTETQINNEL-----TOAQQLTVD 3505
QY 467 QBP--NITLPEK-----EVTFKLDNFRDVEIEKAFGLLYPGWBELEQAR-K 511
Db 3506 KQPLINAKTALQOSLDNQPSITGTEATONTYNAKRQAEQ-----VIQANKIENQAQPS 3561
QY 512 AQASPEKEKSKGLKEFSQ-----QKEENSKAINNOEGLEEDDNTITERLPENSP 562
Db 3562 VQVSDSEKSKVQALSELNNAKSAALRADKQELQQAQYNQ---LIQPTDLNNKKPASITAYN 3618
QY 563 -QYQENAGLGAS-----PDKPYMKDVONQRYLYAKSIOELIKAKDYTKLAKLS 613
Db 3619 QRYQFNSNELSTKNTDRILKEQNPVSADVNN-----ALNKVREV--OOLKNEARALQ 3671
QY 614 NHRTYNISLRLEQDLFVNPRIIPSS-----RDIEKAKFVLDTKTEK 654
Db 3672 NKEDNSALVRAKEQLQQAQVDPSTEGMTQTKDDYNSKQAAQAEISKAQVINDGAT 3731
QY 655 KYWQIYSSASPVFONKWSLFGYRYLLGLDPKQTHIELVKLGKAGLOFEGYENLPSPDN 714
Db 3732 T--QOISNAKTNVERALEALNNAKTGLRADKEELQNAVYNLTQ---NIDTSGKTPASIR 3785
QY 715 LEDLKNIRIKTLPFSQKONFKLSLLDFNNYD-----GEIKA--PEFGLPLFLPKELRR 766
Db 3786 KYNEAKSRIQTOIDSAKNEANSILTDNPNQVSVQVTAALNKIKAVQFELDKATLAKNKEN 3845
QY 767 NSNSGGSGQ-----NSNSPME---QEIIISQFQDNLSNQDLAQFSTKIWEKII--GDE 815
Db 3846 NNALVQAKQOQLOQIVNEVDPTQGMTTDTANNYKSKKRAEDEIQKA-----QQIINNGDA 3900
QY 816 NE---FDQNNRLQYKLLKDLQBSWINKTRDNLWYLYGDKLVKPK-----NNLEAKFKQ 867
Db 3901 TEQOITNETNRVNOAI-----NAINKAKNDL---RADKSOLENAYNQLIQNVDTNGKK 3950
QY 868 ISNLOELLTAFYTSAAALSNNWYODSGAKSTIIFEETAEPLDPKVEKVA---DVOYL 923
Db 3951 PASIOOYQAA---ROAIEQYN---NAKSE-AHQILENSNPVSNEVAQALQKVEAVOL 4001
QY 924 KPHYAIGFDDNAGKFNQEVIRSSRTIYLTSGKSKLEADTTIDQLN-----QAVK 973
Db 4002 KVNDAILHQLN---KENNSALVTAKNLOQSVNDQPLTTGMTQDSINNYEAKNEAQAIR 4059
QY 974 NAPLGLQ-----SFYLDTERFVGFOKLATSLAVQHKQKEKTPKKNLNDGYTLIHDKLLK 1028
Db 4060 NAEAVINNGDATAKQISDEKSKVEQALA-----HLNDAKQQLTADTTTEL 4103
QY 1029 PVIPOISSPEKDFEGKLNQNGQSONV---NVSTFGSTIESPYFTNPFQEDADLDQDQ 1085
Db 4104 QTAVQ-----QLNRGDTNNKPRISINAYNKAIQSL----- 4134
QY 1086 DDSROGNNSLNQEAAGLKKLAILLGNQFIQYQOQNDKEIFEIINVEKVE-----L 1139
Db 4135 --ETQITSAKDNANAVIQPIRTVQEVNNALQVNNALQQLTEAINQLPLSNNDALKAA 4192
QY 1140 SRVFEFLAKTLEDNGKIRVLSDEMSLI VNTTIETKPEMSAVPEVPTKVEQYDPT 1199
Db 4193 RLNLKINKQTVQTDGMT-----QOSIBAYQNAKVAQNESNT 4230
QY 1200 PLAAKTKFVLKFKDQIPVDGSGNISDKWLASIPLVHQMLRLSPVVK--TIRELGLKTE 1257
Db 4231 ALA-----LINNGDAEQOITTEDRVNQOQTNLTQAINGLTVNKEPLET- 4275
QY 1258 QQQQQQQQQQQQPOKKAVRKEELETYNPKDEFNINPLTKAHLTLTSLNVLNNDPNYKI 1317
Db 4276 -AKTALQNNIDQVPSTDGM-TQOSVANYNQLQ-----IAKNEINTINNVLANNP----- 4323
QY 1318 EDLKVIK-NEAGDHOQLAFSL-RANN-----IKRLMTPPI---TFAD---YNPFF 1358
Db 4324 -DVNAIKTKAEAEIRISNDLTQAKNQLVDTQPLEKIKRQLODEIDQGTNTDGMTQDSVD 4382
QY 1359 YNEDWRS-----IDKYLNNKGVSSHQOQAAGNOGSGLIQ----- 1395

4383	NYNDSLSAAIEKGVKNLKRNPVTVEOVKESVANAQQ----	VIQDLQARTSLVPDKTQL	4439	
1396	-----RLNKNIKPTFTPALIALKDRNNTNLSYSDKIIMKPKYLVERSIGVPMWSTGLD	1450		
4440	QEAQNRLENSINQOTDTG-----	TODSLANNYDK--LAKARQNLKTI-----SKVLG	4486	
1451	GYIGSEQTKDGTSSSQKGFQDQFIALGLK-----	NTEYH-----GK	1489	
4487	GOPTVAEIRQNTDEANAHK-----	QALDTARSQTLNREPYNHNNESHNLNAQKDN	4539	
1490	LGLSIRIFDPGNEIAKIK-----	DASNKKGBEKL-----KSYD-	1523	
4540	FKAQVSNAPNHTLETIKKADTLNQSWTALSSESIADYENQKQOQENYLDASNNKRQDYDN	4599		
1524	---LFRKNYLE-----	YEKKSPIAKGWTNIHPDQKEYPNPKLPENYLNLV--	1568	
4600	AVNAAGTLNQPTMSADVIDOKAEDVKRTKTALDGNORLEVAQOQAL--	NHLNTLND	4657	
1569	LNQPKWTLNNSDFITNLFVEPEGSDRGSGTKLKQVIQKQVNNYAD--	WGSAYLTFWY	1626	
4658	LNDAQROTLTTHNSPNSVNOAKEKANTVNTAMTQTKOTIANYDDELHDGNYINADK	4717		
1627	DK-----NIITNOPNVIATIANADVFIKDVKELENDTKLIAPNITQWPNISGSKERFYKP	1681		
4718	DKDAYNNAVNAQNLINQSDANQAQDPAEINKVTQRVNTTKNDLNGNDKLAERKRDAN	4777		
1682	TVPFG-----NWNEN-----	SSMNSQAQPTTWEKIREGPAQALKSSFD-----Q	1722	
4778	TTIDGLTYLNEAQRNKAQENYKASTKTNITSQLDQYNQL--	NIAMQALRNSVNDVNVK	4835	
1723	KTRFVLTTNAPLWKGYP-----	LGFGNGP-----NFKTQDWRLVFQNDNQIAA	1769	
4836	ANSYNEDNGPKYANQAVTHAOTLINAQSNPEMSRDVNVQKTAQVNTAHQNLHGQOKL	4895		
1770	LRVQEQ-----	DRPEKSSDKQKWKIKFKVVIPEEMFNSGNIR	1808	
4896	EQAOSSANTEIGNLPNLTNTQKAKEK-----	ELVNSKQTR	4930	
RESULT 8				
US-09-949-016-11433				
; Sequence 11433, Application US/09949016				
; Patent No. 6812339				
; GENERAL INFORMATION:				
; APPLICANT: VENTER, J. Craig et al.				
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED				
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF				
; FILE REFERENCE: CL001307				
; CURRENT APPLICATION NUMBER: US/09/949,016				
; CURRENT FILING DATE: 2000-04-14				
; PRIOR APPLICATION NUMBER: 60/241,755				
; PRIOR FILING DATE: 2000-10-20				
; PRIOR APPLICATION NUMBER: 60/237,768				
; PRIOR FILING DATE: 2000-10-03				
; PRIOR APPLICATION NUMBER: 60/231,498				
; PRIOR FILING DATE: 2000-09-08				
; NUMBER OF SEQ ID NOS: 207012				
; SOFTWARE: FastSeq for Windows Version 4.0				
; SEQ ID NO 11433				
; LENGTH: 2733				
; TYPE: PRT				
; ORGANISM: Human				
US-09-949-016-11433				
Query Match				
Best Local Similarity 2.8%; Score 271; DB 4; Length 2733;				
Matches 308; Conservative 246; Mismatches 507; Indels 578; Gaps 76;				
Qy	106	IPKQYQVJISEPDDKERFRLGFLHKEJEDGNTAQATKFIYLLPLDMPKALGQVS	165	
Db	1212	LSKKQFSLMSEKDSLEEVQ-----DLKHQIE-GNVSKQAN-----	LEATEKHQDNQTN	1258
Qy	166	Y-----	IVDKNFNNLIHPLSNFSAQSIKPLATLRSDDFTAKLNQNFNQDELWV	214

Db 2214 MSSLQNSRDHANBELDKKRYDASLKELAQKEQGL-----LNRE-----RDALLSETAFS 2265
Qy 1034 ISSPERKDWPE-GKLNQOSQNVNVSTFGSIIESPY-----FS---TNFQEDAD----- 1079
Db 2266 MNSTEENSLHLEKLNQALLSKDBQLLHLSQLEDVYNQVQSFKAWASIQNERDHLWNE 2325
Qy 1080 -----LDQCGDDDS-----ROGNSLONQEAAGLLKQKLAILLGNQFTQYYQ 1120
Db 2326 LEKFRKSEEGKORSAAQFSTPAEVQSLKAMSLQNDRLLE-----LKNLQOQYLYQ 2380
Qy 1121 QN-----DKEIEFEIINVE-----KVSELSF 1141
Db 2381 INQEIITELHPLKAQLEYQDKTAKFQIMQBELQENLSWOHELHQLRMKSSWEIHERRM 2440
Qy 1142 RVBFKLAKLTEDNG-----KTIRVLSDETMSLIVNTTIEKTPEMSAVPE-----VFD 1188
Db 2441 KEQYLMAISDKDQQLSHLQNLRELSSSSQTPKLVQVQROQASPETSPDGSQNLVYE 2500
Qy 1189 TKWVEQVDPPTPLAAKTKFVLKFK-----DQIPVDGSGNISDKWLASIPLVTHQMLRLSPV 1245
Db 2501 TELL-----RTQLNDSLKEIHQELRIQQLNSNFSQLBEEKNTLSIQLCDTSQSLR----- 2551
Qy 1246 VKTIRELGLKTEQQ-----QOQOQOQOQOQ-----QKKAV-RKEEE 1281
Db 2552 -----ENQOHVGDLLNHCAVLEKQVQLQAGPLNIDVAPGAPQEKGVHRKSDP 2600
Qy 1282 LETYNPKDEFNILNPLTKAHLRLTSLNLVNDPNYKIEDLVKINEAGDHL-----AFSLR 1337
Db 2601 EELREPOQSFSEAQ-----QQLCNT--RQEVNELRKLLEERDQORVAENALSA 2648
Qy 1338 ANNIKRL-----MNTPI 1349
Db 2649 EQQIRLEHSEWDSSRTPI 2667

RESULT 9

US-09-710-279-2964
; Sequence 2964, Application US/09710279
; Patent No. 6703492
; GENERAL INFORMATION:
; APPLICANT: KIMBERLY, WILLIAM JOHN
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
; FILE REFERENCE: F034800S
; CURRENT APPLICATION NUMBER: US/09/710,279
; CURRENT FILING DATE: 2000-11-09
; PRIOR APPLICATION NUMBER: 60/164,258
; PRIOR FILING DATE: 1999-11-09
; NUMBER OF SEQ ID NOS: 4472
; SOFTWARE: PatentIn ver. 2.1
; SEQ ID NO 2964
; LENGTH: 5024
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic
; OTHER INFORMATION: amino acid sequence
; FEATURE:
; NAME/KEY: MOD RES
; LOCATION: (5024)
; OTHER INFORMATION: variable amino acid
US-09-710-279-2964

Query Match 2.7%; Score 262.5; DB 4; Length 5024;
Best Local Similarity 19.6%; Pred. No. 3.4e-08;
Matches 392; Conservative 272; Mismatches 731; Indels 605; Gaps 97;
Qy 64 SDYQSV--KKALLNGKTFDPKSEFTDFVSKPFLTNNGRTVLE----IPKQVVISSEFS 118
Db 135 SDNESVNESNYINAE--PEKQAFPTALNNAKEIVNEQOATLDANSINQKAAQILTTKN 192
Qy 119 PEDDKERFLRG-----PHLKEKLEDG-----NIAQSAKTFIYLLPLDMPKALG 162
Db 193 ALDGEQLRAKENAQOEINTNLQTLDAQNSEKGLVNSQRTVEVASQI-----AKAK 246

Qy 163 QYSYIVDNFNFNLIHPLSNFSAQSIKPLALTRSSDFIAKLNQFNQDELWVYLEKFFDL 222
Db 247 ELNKVMEQ-LNNLI-----NGKNQMI-----NSSKFI---NEDANQOQ--AYSNAIASA 289
Qy 223 EALK---ANIRLOTADPFSEKGNLVDPPFVYSFIRPQNO-----KEMA-----SDLN 266
Db 290 EYLKNSQNPDELKVTIEQAINN-----INAINNLNGEAKLTAKEDAVASINNLSGLT 344
Qy 267 OQOKTV--RLYLRTRFSPQAKTILKDYKYKDTFFLSSIDL---KASNTGSLPANENDL- 319
Db 345 NEQKTKENQAVNGSQTRQDVANVLRSKALDOSMTLRDLVNQNVIHSTSYNFNEDSTQ 404
Qy 320 KDQLVDLDDVDYFGGQ-----SETITS-NSQVKVPASERSLKDR----- 360
Db 405 KNTYDAIDNGSTYITQGHNSLNKSTIDQTSIQINTAKNDLHGAELQDRKGTANQEI 464
Qy 361 -VKFKDQOKPRIKFSLEYEDALSFSQJELVSKPNSIKOLVNTATLAR-NLRFSLGKY 418
Db 465 QLGYLNDPQKSAEESLVNGSNTRSEVEHLNEAKSLNNAMKQLRDKVAEKTIVKQS---S 521
Qy 419 NFLFDDLASHLDYFVLVSKAK-----TKQSSITKLFIELPIKISLASS-ILGQ 467
Db 522 DYINDSTEHQRYDQALQEAENIINEIGNPTLNKSEIEQKLOQLTDAQNALQSHLLEA 581
Qy 468 EPNIKLFEKEVYTKLDMFRDVEIEKAFGLL-----YPCVNEEL----- 506
Db 582 KKNAIT---EIN-KLTALNDAQRKATENVQAOQTIPAVNQQLTLDRREINTAMQALRDK 636
Qy 507 --EQARKAQRASF--EKEKSKG-----LKEFSQOKEENSKAIN----- 541
Db 637 VQOQNVHQQSNYPNEDEQPKHYNDSVQAGOTIIDKLQDPIMNKNEIEQAINQINTTQT 696
Qy 542 ---NOEGLEEDDNITERLPEN--SPIQYOQENAGLGASDPKPYMIKDVONQRYVLAKSQIQ 597
Db 697 ALSGENKLUHDOESTNRQIEGLSSLNTAQINAE-----KDLVNQ----- 735
Qy 598 ELIKADYTKLAKLSNRHTYNIISLRKEQLPDVNPRISSRDIKAKFVL--DKTEKNK 655
Db 736 ---AKTRTDVAQKATAKEINSAMS-----NLRDGIQNKEDIKRSSAYINADPTKVTA 785
Qy 656 YKQIYSSASPQNKWSLFGYYRYLLGLDP-----KQTHIELVKLGOKAGLQFEGVENLP 710
Db 786 YDQALQNAENI-----INATPNVELNKATIEQALSrvQQAQDLQDGVQQA 831
Qy 711 SDFNLEDLKNIRIKTLPFSQDNFKLSLLDPNNYVDGEIKAPEGLPLPLPKELRRNSN 770
Db 832 -----NAKQOATQTVNGLNSLNDGQKR--ELNLLI-----NSAN 863
Qy 771 S-----GGSQNSNPWE-----QEIIGFDQNLSNQDLAQFSTKIWEKIIGDEN 816
Db 864 TRTKVQEELNKATESNHAMEALRNSVQNVVDVKQSNVYVNEDEQEHN---YDNVNEAQ 920
Qy 817 EFDQNR--LOYKLLKLOESMINTRDNLWYTLGDKLVKPKNNLEAKFQISNLQE- 873
Db 921 ATINNNAQFVLDKLAIERLTQTVNTTYQDALH-----GTOKLIQDQQAATGIRGLTSLNRP 976
Qy 874 -----LLTAFYTSAAALSNNWNYQDSGAKSTIIFEIATELD-----PKVKEK--V 916
Db 977 QKNAEVAKVTAATTTRDEVN-----IRQEATTLDTAMLGLRKSIKDKNDYTK 1022
Qy 917 GADVYQLKPHYAIGFDDNAGKFNQBVIRSSSTIYLTSGKSKLBADTIDQLNQAVKNAP 976
Db 1023 NSSKYINEDHQDQOQAYDNVAVNNAQHVIDEQTAT-----LSSDTINQLANAVTQAK 1072
Qy 977 LGLQSFYLDTERFGVFGKLAISLAVQHKQKEXTLPKLLNNDGYTLIHDK-LKKPVPIS 1035
Db 1073 SNLHG---DTK-----LQHKDKSAKQTTIAQIQ 1096
Qy 1036 SSPKDWKDFEGKLNQNGQSONVNVSTFGSIIESPYSTNFQEDADLDQDQDSDSQGNNSL 1095
Db 1097 -----NLNSAQKHMB-----DSLIDNSTRSTQVQHDLTEAQ-----AL 1129

Db 1729 TPNATEEKQDAIQLNGKREDEKVLNINQDR-----:||:|:|:|:RDNEVEQHKNIGLOE 1774
Qy 825 -----QYKLLKOLQESWINKT---RDNLWYTLGDKLVKPKQNLN- AKFRQJISN 870
Db 1775 LETIHANPRKSDALQELQTRFISQTELINNNKDAT--NEEKDEAKLLBISKNKTTIN 1831
Qy 871 LOELLTAFYTSALNNNNYQDSAKSTIIFEEIABLDPKVEKVGADVYQLKPHYAIG 930
Db 1832 INQAQTNQVNDKNGHNEIATIIIPATTIKTDAKTAIDKBAEQOVTI----- 1879
Qy 931 PFDNAGKNOEVIRSSRTIYKTSKGL- EADTIDOLNOQVKNAPGLQSFYLDTERF 989
Db 1880 INGNNDATDEEKAERKLVKAKIEAKSNITNSDTEREVNGAKTN---GLEKIN----- 1930
Qy 990 GYFQKLA TSLAVQHKQKKTLPKLNNDGYTLIHDKLKPVPQISSSPKDFEGK--- 1046
Db 1931 -----NIQPTSTQTKNAQ-----EINDKAQBOLI-QINNTPDATEEKBQAT 1972
Qy 1047 -----LNQNGOSQVNVSTFGSI-----IESPYFTNFQEDADLDQD 1083
Db 1973 NRVNAGLAQAIQINNAHSTQVNESKTSIATIKSVQPNVKKPTAINSLTQEAANNQKT 2032
Qy 1084 --GODDSRGNNSLNQAG--LLKQKLAILLGNQFIOYQONDKEIEFEIINVEKVSSEL 1139
Db 2033 LIGND-----GNATDDEKEAAKQLVTQKL-----NEQIOKIHESQDQNDQVNVKAQAITAI 2083
Qy 1140 SF-----RVEFKLAKTLENGKT-IRVLSDETMSLTIVNTTIEKTPMSAVPEVFDT- 1189
Db 2084 KLINANAKRQDAIILNTLNAESKSDIRANQDA-----TTEEKN---TAIQSIDDTL 2133
Qy 1190 -----KWVEQYDPTPLAAKTKFVLKPKQOIPVDGSGNSIDSKWLA 1229
Db 2134 AQARNNINGANTNALVDENLEDGKQKQRLVLSQTQKQK-----A 2175
Qy 1230 SPLVTHQOMLRSLPVKVTIRELGLKTEQOQOQOQOQOQKKA VKKEBELETYNPKD 1289
Db 2176 DIAQAIGQOR-----STIDQONATTEEKEQALERINQETNGVNDRIQAALANQWTD 2228
Qy 1290 EFNIL-----NPLTKAHLRLTSLNVDNPNYKIED-----LKVKIN 1325
Db 2229 EKNNILETIRNVEPIVIVPKANEIRKKAABQTTLINQONQATLEEQIALGKLEEVN 2288
Qy 1326 EAGDQLAPSLRANNIKRLMNT-----PITFADVNPFFYVNEHWRS-IDKYLNNKG 1375
Db 2289 EA-LNQVSAHNSNDVKTAENNGIAKISEVHPETIIKRNAKQIEBQDAQSQIDTI--NAN 2345
Qy 1376 NVSSHQOQAGGQSGGLIQRNL-----KNIKPETFTPALIALKDRNNTLSNVSKI 1428
Db 2346 NKSTNEEKSA-----IDRVNVAKIDAINNITNATTIQLVNDKNGSNTSIS----- 2392
Qy 1429 IMIKPKYLVERSIGVPWSTGLDGYIGSE-----QTKDGTSSSQOQKGFQDQFIQ-- 1477
Db 2393 -QILPSTAVK-----TNALAALASEAKNNAIIDQTPNATAEKESEANNKVDRLQEE 2443
Qy 1478 -----ALGLKNTEYHGKLGLSIRIF---DPGNELAKIKDASNKKGE----- 1515
Db 2444 ADANILKAHTTDEVNNIKNAQVQINAVQVEIKQNVKNQNLQNFIDNQKIIIENTPDAT 2503
Qy 1516 -EKLKSYDLFKNYLN-----EYEEKSPKIAKGTWNIHPDQ 1550
Db 2504 LEEKAEANRLLQNLVLTSTDEIANVDHNNNEVDQALDKAPKTEEIIVPQVSKRDLVNAIQ 2563
Qy 1551 KVPNPNQKL PENY-----LNLVLNQPKVTL-----YNSDPI 1584
Db 2564 EAPNSQTQIENQOEATNEEKEALNKINQLNQA-KVNIDQAQSKVDSDAKTSIQDI 2622
Qy 1585 TNLVFEPEGS DRG-----SCTKLKQVIQK----- 1608
Db 2623 EQIQPHQPKATGRHRLNEKANQOQSTIATHPNSTIEERQEAQKLEVLKKAIAIDKG 2682
Qy 1609 QVNNYVADGSAVLTFTWYDKNI-----ITNOPNVITANI-----ADVFI-KVYKELEDNT- 1657

Db 2683 QTND-----VEKTVVNGIAEINILPATTVKDKAKADVNNAEKEQKNLQINSN 2730
Qy 1658 -----KLI-----APNITQWPNISGSKEK-----FYK 1680
Db 2731 DEATTEBKJVASDNLHNHVETTNQATBEDAPDTNQ-----VNVKKNKGIGTIRDIQPLVVKK 2786
Qy 1681 PTV---PFGWENENSSMSQAOTPTWEKIREGFAQALKSSFDQ-----KTRTFVLTT 1731
Db 2787 PTAKSKIESAVEKKKTEIN-OTQNAHDEVRG--LNQLNQIHEKAKNDVNSQTNQOVE 2843
Qy 1732 NAPLPMKYGLFQGNFKNFTQDRLVFNQNDNQIAALRVQBQDR-----PEKSSDKD 1786
Db 2844 NAE-----QNSLD-QINNFRPDFSKRNAVAEIVKAQOQNKIDEIEQEFSAQEE 2891
Qy 1787 KQWIKKPVVPEMFNS-----GNIRFVGWQIQGPNLWLPVIN----- 1827
Db 2892 KDNALQHLDEQVKEIININQANTDNEVDNAKTSGLNNI-----TEYRPEYNNKKNAILKL 2947
Qy 1828 -----SSVYDFYRGTDGSDNVAN--LNVAPMQVKT---IAFTNNAFNNVPKFEFNISK 1875
Db 2948 YDVSDDQEAINGPDATEBELQENSKLNKILLDAKQIGLAHTNNEVDIYNE--VSQ 3005
Qy 1876 KI 1877
Db 3006 KM 3007

RESULT 12
US-08-755-587-44
; Sequence 44, Application US/08755587
; Patent No. 6045997
; GENERAL INFORMATION:
; APPLICANT: Futreal, Phillip A
; APPLICANT: Wooster, Richard F
; APPLICANT: Ashworth, Alan
; APPLICANT: Stratton, Michael R
; TITLE OF INVENTION: Materials and methods relating to the
; TITLE OF INVENTION: Identification and sequencing of the BRCA2 cancer
; TITLE OF INVENTION: susceptibility gene and uses thereof.
; NUMBER OF SEQUENCES: 222
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Bell Seltzer Park & Gibson
; STREET: 310 UCB Plaza, 3605 Glenwood Avenue, PO Drawer 31107
; CITY: Raleigh
; STATE: NC
; COUNTRY: USA
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/755,587
; FILING DATE: 25-NOV-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9523959.6
; FILING DATE: 23-NOV-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9525555.0
; FILING DATE: 14-DEC-1995
; APPLICATION NUMBER: GB 9617961.9
; FILING DATE: 28-AUG-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Kenneth D Sibley
; REGISTRATION NUMBER: 31,665
; REFERENCE/DOCKET NUMBER: 5405-135
; INFORMATION FOR SEQ ID NO: 44:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3418 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein

US-08-755-587-44

Query Match	2.6%	Score 256;	DB 3;	Length 3418;
Best Local Similarity	19.2%	Pred. No. 5.3e-08;		
Matches	365;	Conservative 258;	Mismatches 656;	Indels 626; Gaps 92;
QY	85	EFTDPSKFDPLTNGRTVLEIPKQVWISERSPE-----DKERF 126		
DB	1126	EFTQF-RKPSYILQ--KSTFEPENQMTILKTSEECRDADLHVIMNAPSIGQVDSKQF 1182		
QY	127	RLGFHLKEKLEDGNIAGSATKFTYLLPLDMPKALGOYSYIVDKN-----171		
DB	1183	EGTVEIKR-----KFLAGLLKNDCKNSAGS--YLTDENEVGFRGFSANGTKL 1227		
QY	172	-----FNNLIHPLSNFSAQSIKPLALTRSS--DFIAKLQNPNODELWVYL 216		
DB	1228	NVSTEALQAKVLFSD--IENISEETSAEVHPISLSSCKHDSVSMFKIENHNDKTVE 1285		
QY	217	EKFFDLBALKANIRLQTADFSPEKGNLVDPFVYSFTRPNQO-----258		
DB	1286	KNNKQILQNNIEMTT-----GTFVEBITENYKRTENEDNKYTAASRNSHNLFDG 1338		
QY	259	-----KEWASDLNQDKTVRLYLRTFSPOAKTILADYKYKDETRFLSSIDLK- 305		
DB	1339	SDSKNDTVCIHKDETOLLFTDQHNICLKSGFMKEGNTQIKR-DLSDLUTFLVAKAQE 1397		
QY	306	-----ASNGTSLFA--NENDLKQDLQVLDLDVSDYFGGQSETITSSNQVPPASERSLK 358		
DB	1398	ACHGNTSNKEQLTATKTEQNIKD-----FETSD-----TFQATAGKNIISVAKELFN 1444		
QY	359	DRVKFKDQOKPRIEKESL-----YEYDALSPY-----SOL 389		
DB	1445	KIVNF-FDQRPPEELHNFSLNSLSHSDIRKKNMIDILSYEETD1VKHKILKESVVPVGTGNOL 1503		
QY	390	QELVSKENSTKDLVNA7L-----ARNLAFSLGKYNFLFD-----423		
DB	1504	VTQCGQPERDEKIKETPLLGFHTAGSKVKVIAKESLDKVNLFDEKEOGTSEITFSHOW 1563		
QY	424	-----DLASHLDYFVLVSKAKIKQSSITKK-----LFIELPIKISLKSSTLGDQ 467		
DB	1564	AKTLKYREACKDLELACETIETIAAPKCKEQMSLNDKNLVSIETVPPKLLSDNLRCQ 1623		
QY	468	EPNITLFEKEVTFKLDNPROVIEKAFGLLYGVNEE-----LEQAKAQRASFEKS- 522		
DB	1624	TENLKT--SKSIFLKVKHENVEKETAKSPATCYTNQSPYSVIENSALAFYTSCSRKTSV 1681		
QY	523	-----KKGLKE--PSQOKEENSKAINNOGLEEDDNITERLPPNSPIQYQOENAGLG 572		
DB	1682	SQTSLLKAKWLRREGIFDQOPERINTA-----DYGVNLYENNSSTIAEN- 1727		
QY	573	ASPDKPYMIKDVQNYRYLAKSQIQLIKAKDYTKLAKLSNRHTYNI SLRLKEQLFDVN 632		
DB	1728	-----DKNHL--SEKQDTYLSNSS-----MSNSYSYH-----SDEVYN-D 1759		
QY	633	PRIPSSRDEKAFVLDKTEKNKYWQIYSSASVPFQNKWSLFGYRYLLGLDLPQTIHEL 692		
DB	1760	SRILSKNKLDSG---ISPVLKVEDQKNTSFSKVISNVKDANAY-----PQTVNE- 1806		
QY	693	VKLQKAGLQFEGYENLPDFNLDELKNIRIKTLPFSOKDNFKLSLDDENNYVDGEIKAP 752		
DB	1807	-----DICVEEL--VTSSSPCKKNKNAIKLSISNSNFP---EVGPP 1842		
QY	753	EFGLPFLPKELRRNSSGGSQNSNPSWQE:ISQFKD-----QNL5NQDOLA 801		
DB	1843	AFRI-----ASGKIRLCS--HETIKVKVDIDSPSKVIFTKENNENSKIC 1885		
QY	802	QFSTKIW-----EKIIGENEFQNNRLOYKLLKLDQESWINKTRDNLWYTLGD 851		
DB	1886	Q--TKIMAGCYEALDDSEDTLHNSLNDDECSMHSHKVFADIQISEETLQHNQNM--SGLEK 1941		
QY	852	KLKVKPKGN-NLEAKFRQISNLQELLTAFYTSAAALSNNNNYQDSGAKST-----899		
DB	1942	VSKISPCDVSLFSDICKCSIGKL-----HKSVSANTCGIFSTASGKSQVQSDASLQNR 1997		

QY	900	IIFEIEAELDPKVKEKV-----GADVYQLKFHYAIGFDD---NACKFQNEQVIRSSRTI 950		
DB	1998	QVFESEIDSTKQVSKVLFSKNEHSDQLTRENTAIRTPHEHLSQKGFYSNVVNSAFSG 2057		
QY	951	YLKTSCKSLKLEADTTIDQLNQAVKNAPLGLQSF-----YLDTEREGVQKLATSLA 1000		
DB	2058	FETASGK---QVSIILESSLHKVGV---LEEFDLIRTEHSLHYSPTSQNV-----2102		
QY	1001	VQHKKQKETLPKLLNNDGYTLIHDLKPKVP1Q1SSSPEKDFEGKLNQNGOSQNVNST 1060		
DB	2103	-----SKILPRVDKXNPEHCNVSEMEKCSKEFKLS-----NNLNVEGSGSENN---2146		
QY	1061	FSGIIESPVFSNPFQEDADLDQDQDSDSQNGNSLDNQAGLLKOKLAILLNGNQFYQY 1120		
DB	2147	-HSIKVSPYLS-QFOOD-----KQQLVLGTKVSIVENIHVLGKEQ 2184		
QY	1121	QNDKEIEFEIINVEKVSSELSFRVEFKLAKT-----LEDNGKT-IR 1159		
DB	2185	ASPKVVKMBIGKTTFTSDVPVKTNIEVCSTYKDSSENYFETAEVIAKAFMEDDELTD 2244		
QY	1160	VLSDETMSL-----IVNTTIEK---TPMSAVPEVFDTKWVEQYDPRTPLAAKTK 1206		
DB	2245	LPSHATHSLFTCPENEEMVLSNRIGKRGEPILLVGSESIKRNLLNEFDRIIENQEK 2303		
QY	1207	FVLKFKQOIPVDGSGNI SDKWLASIPLVHQMLRLSPV---KTIRELGUKTEQOQOO 1262		
DB	2304	--LKASKSTP---DGTIKDR---RLFMHH--VSLEPITCVPFRTTKE-----2340		
QY	1263	QOQOQOQOPOKKAVERKEELETYNPKDENILNPLTKAH---RLTL-----SNL-VNNDPN 1314		
DB	2341	--RQEIQNPNTA-----PQOEF-----LSKSHLYEHLTLEKSSSLAVSGHPF 2382		
QY	1315	YKIEDLKVIKNEAGDHLAFSLRANNIKRLMNTPIFEADYNPFYFYNNEDWRSIDKYLNN- 1373		
DB	2383	YQVS-----ATRNEKMRHLJITGRPTKVFPVPFKTSHFHRVEQCVRN 2426		
QY	1374	-----KGNVSHQOQAAGNOGSGLIQRLNKNIKPE-----TFTPA-----LIALKDR 1416		
DB	2427	NLEENRQKNIDGHGSDSKKNINDNEIHQFNNNSNQAAVPTTKCEEPDLDI-----2481		
QY	1417	NNTNLSNYSD-KIIMKPK-----YLVERSIGVPMWSTGLDGVIGSEQTKDGT 1463		
DB	2482	--TSIQNARDIQDMRIKKQRORVFPQPSLYLAKTST-LP-RISLKAAGQVQPSACSH 2537		
QY	1464	SSSQOQGFDDFIQALGLKNT-----YHGK---IGLSIRIFDPGNEIAKINDASN 1511		
DB	2538	KQLYTVGVSKHKIK-INSKNAESSFOHTEDYFGKESLWTGKIGQLADGG---WLIPSDNG 2593		
QY	1512	KGGEKLLKSYDLFKNYLNEYEKSPK-IAGWTNIIHPDQKEYPNQKL-----PEN 1563		
DB	2594	KAGKEE-----FYRALCDTFGVDPKLIISHIWWYNH-----YRWIIWKLAAMECAFPKE 2641		
QY	1564	YLNVLNQPWKVTLNYSDFITNLFVEPEGSDRGSGTKLKQVIOQVNNNNYADWSAYLT 1623		
DB	2642	FANRCLUS-PERVLLQIKRYDTEI-----DRSRRSAIKKIMER-----DDTAAKTL 2686		
QY	1624	FWYDKNIITQNPVITANIADVFIKDVKELEDNTKLIAPNITQW 1668		
DB	2687	VLCVSDIIS-----LSANISSETSSNKTSS-ADTKQVAILIELTDGW 2725		

RESULT 13

US-08-328-254-6
; Sequence 6, Application US/08328254
; Patent No. 5710022
; GENERAL INFORMATION:
; APPLICANT: Zhu, Xueliang
; APPLICANT: Lee, Wen-Hwa
; TITLE OF INVENTION: A No. 5710022el Nuclear Mitotic Phosphoprotein
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSES: Campbell and Flores

STREET: 4370 La Jolla Village Drive, Suite 700	
CITY: San Diego	
STATE: California	
COUNTRY: USA	
ZIP: 92122	
COMPUTER READABLE FORM:	
MEDIUM TYPE: Floppy disk	
COMPUTER: IBM PC compatible	
OPERATING SYSTEM: PC-DOS/MS-DOS	
SOFTWARE: Patent in Release #1.0, Version #1.25	
CURRENT APPLICATION DATA:	
APPLICATION NUMBER: US/08/328,254	
FILING DATE: 24-OCT-1994	
CLASSIFICATION: 435	
PRIOR APPLICATION NUMBER: US 08/141,239	
FILING DATE: 22-OCT-1993	
ATTORNEY/AGENT INFORMATION:	
NAME: Campbell, Cathryn A.	
REGISTRATION NUMBER: 31,815	
REFERENCE/DOCKET NUMBER: P-CJ 1191	
TELECOMMUNICATION INFORMATION:	
TELEPHONE: (619) 535-9001	
TELEFAX: (619) 535-8949	
INFORMATION FOR SEQ ID NO: 6:	
SEQUENCE CHARACTERISTICS:	
LENGTH: 2482 amino acids	
TYPE: amino acid	
TOPOLOGY: linear	
MOLECULE TYPE: protein	
US-08-328-254-6	
Query Match 2.6%; Score 255; DB 1; Length 2482;	
Best Local Similarity 18.9%; Pred. No. 3.8e-08;	
Matches 378; Conservative 333; Mismatches 764; Indels 524; Gaps 91;	
QY 61 NPTSDYQSVKALLNGKTFDPKSSFTDFVSKDFL-TNNGRTVLKPKVQVVISBSP 119	
DB 434 NRKNEQLKEAF-----AKEHQBFVLKLAFAERNQMLLETVQOALRSEMTD 484	
QY 120 EDDKERFRLGHLKKEKEDGNIQAQATKFIYLLPLDMPKAAALGOYSYIVDKNFNNLI--- 176	
DB 485 NON-----NGKSAGGLKQE-----IMTLKEQNK-----MQEVDNLLQEN 521	
QY 177 -----IHPLSNPSAQSIKPLALTRSSDFIAKLNQ--FNNQDELWV-----YLEK 218	
DB 522 EQLMKVMTKHCQNLSEPIRNSVKERESE-----RNQCNFKPQMDLEVKEISLDSYNAQ 577	
QY 219 FFDLEALKAN-----IRLOTADPSFEKGNLVD-----PFV 248	
DB 578 LVQLEAMLRLNKLQSESEKECELOHELTQIRGDLTSNLDQMOSQISGLKQCEIDAE 637	
QY 249 YSFIRNPQKQEWASDLNDQKTVRLYLRTFS-----POAKTILKDYKDETFLLSSID 303	
DB 638 EKIYSGPHE-----LSTSQNDNAHLQCSLQTTWNKLNLEKEICEILQAEKVELVTELNDSR 693	
QY 304 LKASNGTSLFANE-----NDLKQDLVDLLDVSYPGGO-SETITSNSQKVPV 351	
DB 694 SECITATRWAEVKGKLLNEVKILNDDSGLLHGLL--VEDIPGGEFGEPQNEQHPVSLAP 751	
QY 352 ASERSLKDRVKFKDOOKPR-----EKFSLEYDALSFSYQQLVSKPNSIKDLVNATL 407	
DB 752 LDESNSYELHTLSDKEVQMHFAELOEKFSLQSEHKILHDQHCQSSKMSSELQTTVDSLK 811	
QY 408 ARNLRLSLGKYNF--LFPDLASHLDYFEL--VSKAKIKQSSITTKL-----FIELP 454	
DB 812 AENLVLTNLRNFOGDLVKEMQLGEEGLVPSLSSCVPDSSLSLSDGSSFYRLLEQT 871	
QY 455 IKISLKSSILGQEPN-----IKTLFEKVTFKLDNFRDVEIEKAPGLLYPGVNEEL 506	
DB 872 GDMSSLNLEGAVSANQCSVDEVFCSSLQEENLTRK-----ETPSAPAKGV-----EEL 920	
QY 507 EQARKAQRASFKEKSKGLKEFSQOKE--ENSKAINNQGLEDDDNITERLPENSPIQVQ 565	
DB 921 ESLSCEVYVRSQLEKLEKESQGIKMKETIQLEQLLSERQELDLCKRKQVLSENE--OMQ 978	
QY 566 QENAGLAGSPDKPYMIKDVQNRYYL---AKSQIQELIKAKDYTKLAKL-----611	
DB 979 QKLTSTVLEMSKLAABKQTEQOLSLEVARLOQL-----DLSSRLSLGIDTEDIAQG 1034	
QY 612 -----LSNRHTYNISLRKE---QLFDVNPRIPISSRDIEKAKFVLDKTKKNYMQIYS 661	
DB 1035 RNESCDISKETSETTERPKHDVHQICDKAODLNLDEK---ITETGALAKPTGCSG 1091	
QY 662 SASPFQNKWSLFGYYRYLLGLDPKQTIHBLVKLGKAGLQFEGYENL-PSDF--NUEDL 718	
DB 1092 EQSPDTNYEPP-----GEDKTQGSSECI-----SELSFGPNALVPMDFLGNQEDI 1137	
QY 719 KNTRIKTLPFSQONFKLSLLDFNNYVDGIEKAPFGLPLFKELRRNSSSGSQNSN 778	
DB 1138 HNLQLRV---KETSNNENLRLLHVIEDRDRK-----1165	
QY 779 SPWEOETISQFQDONLSNQDLAQFSTKI-----WEKIIGD---ENEFQDNRRLOQY--- 826	
DB 1166 -----ESLLNEMKELDSKLHLQEVQLMTKIEACIELEKIVGELKKENS-DLSEKLEYSCD 1220	
QY 827 --KLLK-----DLOESWINKTRDNLVYTLGDKLVKPKNNLEAKPRQISNLOEL 874	
DB 1221 HQELLQRVETSEGLNSDLEMHADKSSRED-----IGDNV-AKVNDWSKBRFLDVEN--EL 1272	
QY 875 LTAFYTSAAALSNNWNYQDSGAKSTIIFESIAELDPKVKKEGVADVVQLKAFHYAIGPDDN 934	
DB 1273 SRIRSEKASIEHEALYLE-----ADLEVVOQTEKLCLE-----KDN 1307	
QY 935 AGFNQEVIRSSRTIYKTSKSKL--EADT-----IDQLNQAVNAPLAGLSFYLD 985	
DB 1308 ENK--QKVIIVCLSEELSVTISERNQRLGELDTWSKTTALDQSEKMKETQELSEHSE 1365	
QY 986 TERFGVPKQLATSLAVQHKQKKT-LPKKLNNDGYTLIHDKLKPVPITPQISSSEPKD--- 1041	
DB 1366 C-----LHCIQVAEAEVKEKTELLQTLSSDVELLKQKTH---LQEKLSLEKDSQA 1414	
QY 1042 -----WFEKLNQNGSQSNVNV---STFGSIIIESPVFSNTFOE--DADLDQDG--- 1084	
DB 1415 LSLTKCELENQIAQANKEKELLVKESLQARSESDEYKLVNSKALVKAALVKEGFAIR 1474	
QY 1085 -----QDSSRQNNSLNDQEAEL-LKQKLAITLLGNQFIQYVQONDKEIFEIINVEKVEL 1139	
DB 1475 LSSTQEEVHQLRGIEKLRVRIEADKQLHIAEKLEREREND-SLKDKVENLERELQOM 1533	
QY 1140 SFRVEFKLAKTLEDNGKTIIRVLSDETMSLIWNTTIETKTPMSAVPEVPTDKWBOYDPR 1199	
DB 1534 S-----EENQELV-ILDAENSKAEVETLTKQIEEMARSLKVPFELDVLTLRSEKE 1581	
QY 1200 PLAAKTFVLKFKDQIIPVDGSGNIS--DKMLASIPVLIHQOMLRSLPVVKTIRRELGLKTE 1257	
DB 1582 NLTRQIQ-----EKQQLSELDKLSSFKSLLE-----KEQAEIQIKKE 1621	
QY 1258 QQQQQQQQQQQQQQPKKAVRK---EDEL-----ETYNP--KDEFNINPLTKAHLRLTSLN 1307	
DB 1622 SKTAVEMLQNLKELNEVAALCGDQELMKATEQSLDPPPIEEHQLRNSIEK-----LRA 1676	
QY 1308 LVNNDPNYKIBDLKVIKNEAGDHQALFSLRANNIKRLMNTPIITADYNPPFYNYEDWRSI 1367	
DB 1677 RLEADEKQKQVQLK-ESBHADLLKGRVENLERELEIART-----NQEHAAL 1725	
QY 1368 KYLNKNGVSHQOQAAGNQG-SGL-IQRLNKNIKETPTTALIALKDR-NNTNLGNY 1424	
DB 1726 EAE-NSKGEVTLKAKTEGWTQSURGLDELDVVTIRSEKENUTNELQKEQERISELEIINS 1784	
QY 1425 S-DKTIIMTKPKYLVERSIGVPWSTGLDGYIGSEQTQKGTSSSSQKGFQDPQFIQALGLKN 1483	
DB 1785 SFENILOKEQEKV-----QMKESSTAMEML---QTKLKNLERNV 1822	
QY 1484 TEYHG-----KLGLSIRIFDPGNELAKIKDASNKYGEEKLLKSYDLFKNYLN----- 1530	

Db	1843	AFRI-----ASGKIVCVS---HETIKKVKIDFTDSFSKVIKENNENSKIC	1885
Qy	802	QFSTKIW-----EKIIGDNEFDQNNRLQYKLLKLOESWINKTRDNLWTVLGD	851
Db	1886	Q--TKIWAGCYEALDDSDGILHNSLDNDECSTHSHKVPADIQSEILQHNQM--SGLEK	1941
Qy	852	KLKVKPN-NLEAFRQISNLOELLTAPYTSAAASNWNYYQDSGAKST-----	899
Db	1942	VSKISPCDVSLETSIDCKCSIGKL-----HKSVSANTCGIFSTASGSKVQSDASLQAR	1997
Qy	900	IIFEIABLDPKVKEK-----GADVYQLKFHYAIGFDD---NAGKFNQEVIRSSRTI	950
Db	1998	QVFSEIEDSTQVFSKVLFKSNEHSDQLTRENTAIRTPHELISQKGFYSNVNVSASFSG	2057
Qy	951	YLKTSGSKLEADTIDQLNQAVKAPLGLQSF-----YLDTERFGVQKLAISLA	1000
Db	2058	FSTASGK---QVSILESSLHKVKG---LEBFDLIRTEHSLHYSPTSRQNV-----	2102
Qy	1001	VOHKQKEKTLPKLNNNGYTLIHDKLKKPVIPOISSPEKDPFGKLNQSGSQNVNVT	1060
Db	2103	-----SKILPRVDKRNPEHCNVSEMEXTCSKFKLS-----NNLVNVEGGSENN---	2146
Qy	1061	FGSITESPYSTNFOEDADLQDGDQDSDRQGNNSLDNOEAGLLKQKLAAILGNFIQYQ	1120
Db	2147	--HSIKVSPYLS-QFOOD-----KQQLVLGTVKUSLVENIHLVGLKEQ	2184
Qy	1121	QNDKEIEPEIINVEKVSLSFRVEFKLAKT-----LEDNGKT-IR	1159
Db	2185	ASPKNVKMEIGTETFSVPVKTNIEVCSTYKDSSENYFETEAVEIAKAFMEDDELTDK	2244
Qy	1160	VLSDETMSL-----IVNTTIEK---TPMSAVPEVFDTKWVEQYDPRTPLAATK	1206
Db	2245	LPSHATHSLFTCPENEEMVLNSRIGKRGREPLILVGEPISIKRNLNLEFDRIIENQES-	2303
Qy	1207	FVLKFKDQIPVDGSGNIIDKWLASPLVIHQMLRLSPW---KTIRELGLKTEQOQQQ	1262
Db	2304	--LKASKSTP---DGTIKDR-----RLFWMH--VSLEPITCVPPRTTKE-----	2340
Qy	1263	QOQQOQQOQPKAKVAKREBELETYNPKDEFNINPLTKAH---RLTL-----SNL-VNNDPN	1314
Db	2341	--RQBIQNPNFTA-----PGQEF-----LSKSHLYEHLTLEKSSSNLAVSGHPF	2382
Qy	1315	YKIDELKVIKNEAGDHLAFSLRANNIKELMNTPTITPADYNPFFYNNEDWRSIDKLYLN-	1373
Db	2383	YQVS-----ATRNKMRHLITTGRTKTVFVPPFKTSHFHRVQCVRN	2426
Qy	1374	-----KGNVSSHQOAGGQSGGLIQRNLNKNIKPE---TFTPA-----LIALKDR	1416
Db	2427	NLEBRQKQNIHGHSDDSKKNINDNEIHQFNKNSNQAAAVTFTKCEEPDLDI-----	2481
Qy	1417	NNTNLSNYSK-KIIMIKPK-----YLVERSIGVPMSTGLDGYIGSEQTKDGT	1463
Db	2482	--TSLQNARDIQDMRIKKQRQRPQPSGLYLAKTST-LP-RISLKAAGVQVPSACSH	2537
Qy	1464	SSSQKQGFQDFQIALGLKXTE-----YHGK-----LGLSIRIFDPCNELAKIKDASN	1511
Db	2538	KOLYTVGVSKHCIK-INSKNAESFQFHTEDYFGKESLMTGKGIQLADGG---WLPISNDG	2593
Qy	1512	KGEEKLLKSYDLFKNYLNEYEKKSPK-IAGQWTHIHDPQKEYPNPNOKI-----PEN	1563
Db	2594	KAGKEB-----FYALCOTFGVDPKLISRIVVYNH-----YRWI IWLAAWECAFPKE	2641
Qy	1564	YLNVLNLQPKWTLNSSDFITNLFVEPEGSDRGSGTKLKQVIOKQVNNNAYDWGSAYLT	1623
Db	2642	PANRCLS-PERVLQKLYRYDTEI-----DRSRRSAIKKIMER-----DDTAAKTL	2686
Qy	1624	FWYDKNIINQPNVITANTADVFIDKVELENTKLIAPNIQW	1668
Db	2687	VLCVSDIIS-----LSANISSETSSNKTSS-ADTKVAIIELTDGW	2725
RESULT 15			
US-09-044-946-2			

Sequence 2, Application US/09044946																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																						</
-------------------------------------	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	----

Db 1286 KNNKQLILQNNIEMTT-----GTFVEBITENYKNTENEDNKYTAARNQHNLLEFDG 1338
Qy 259 -----KEMASDLNQDQKTVRLYLRTFFSPQAKTILDKYKQDFTLSSIDLK- 305
Db 1339 SDSSKNDTVCIHKDETDLLFTDQHNICLSLGGQFMKEGNTQIKE-DLSDLTJFLEVAKAQE 1397
Qy 306 -----ASNGTSLFA--NENDLKOOLDVDLLDSDYFGQSQETITSSNQKVPVPSERSLK 358
Db 1398 ACHGNTSNKQLTATKTEQNIKD-----FETSD-----TFFQASGNKISVAKESFN 1444
Qy 359 DRYKFKKDOQKPIREKFSL-----YEXDALSFY-----SOL 389
Db 1445 KIVNF-PDQKPELHNFSLNSLHSDIRKKNMILSYEETDVIKHKILKESVPVGTGNQL 1503
Qy 390 QELVSPNSIKDLVNATL-----ARNLRFSLGKYNFLFD-----423
Db 1504 VTFQGPDERDEKIEPTLLGFHTASGKVKIAKESLDKVNLFDEKEQGTSEITTSFSHOW 1563
Qy 424 -----DLASHDYFVLVSKAKIKQSSITKK-----LFIEPLIKISLSSILGDO 467
Db 1564 AKTLKYREACKOLELACETIETAAAPCKEMQNSLNDKMLSVIETVPPKLLSDNLCRQ 1623
Qy 468 EPNIKTLFEKEVTFKLDNFRDVELEKAFGLLYPGVNEE-----LEQARKAQRASFEKEKS- 522
Db 1624 TENLKT--SKSIFLKVKHENVKEKTRAKSPATCVTNQSPYSVITENGALAPYTCSCRKTSV 1681
Qy 523 -----KKGLKE--FSQOKEENSKAINNQEGLLEDDNITERLSPNSPTIOYQENAGLG 572
Db 1682 SQTSLLEAKKWLREGFDGPERINTA-----DYVGNVLYENNSNSTIAEN---1727
Qy 573 ASDPKPMI KDQVQRYLAKSOIELIKAKDYTKLAKLSNRHTYNIURLKELQFDPVN 632
Db 1728 ---DKNHL---SEKQDTYLSNSS-----MSNSYSYH-----SDEVYNS 1760
Qy 633 PRIPSSRDIEKAKFVLDDKTEKNYQWQIYSSASPVFQNKWSLFGVYRLLGLDLPKQTHIEL 692
Db 1761 GYLSKNKLDGIEPVLKNVEDQK---NTSFSKVISNVKDANAY-----PQVNE- 1806
Qy 693 VKLGQKAGLOFEGYENLPDFNLEDLKNIRIKTFLFSQKDNFKLSLLDFNNYDGEIKAP 752
Db 1807 -----DICVEEL--VTSSSPCKNKNAAIKLSISNSNF---EVGPP 1842
Qy 753 EFGPLPLPKELNRSSNSGSGNSNSPWEQELISQPKD-----QNLNQDOLA 801
Db 1843 AFRI-----ASGKIVCVS---HETIKVKYKIDFTDSFSKVIKENNENSKIC 1885
Qy 802 QFSTKIW-----EKIIGBENFDQNNRLQYKLLKDQESWINKTRDNLNLYWTVLGD 851
Db 1886 Q--TKIWAGCYEALDSEDILHNSLDNDECSTHSHKVFADIQSEELQHNQNM--SGLEK 1941
Qy 852 KLVKPKN-NLEAKFRQISNLQELLTAFTYTSAAALSNNWNYQDSGAKST-----899
Db 1942 VSKISPCDVSELTSDICKCSIGKL---HKSVSANTCGIFSTASGSKVQVSDASLQNR 1997
Qy 900 IIFEETAELDPKVEKY-----GADVYQLKFHYAIGFDD---NACKFNOEVRSSRRTI 950
Db 1998 QVFSEIEDSTKQVSKVLFKSNHSQDLTREENTAIRTPHEHLISQKGFSNVNVNSAFSG 2057
Qy 951 YLKTSGSKLEADTIDQLNOAVKNAPLGLQSF-----YLDTERFCVFOKLATSLA 1000
Db 2058 FSTASGK---QVISELSLHKVGV---LEEFDLIRTEHSLHYSPTRSQRNV-----2102
Qy 1001 VOHQKEKTLPKKLNNNDGYTLIHDLKKPVIPOI SSSPEKDMFEGKLNQNGQSONVNST 1060
Db 2103 -----SKILPRVDKRNPRHCNVSEMEKTSCKEFLS-----NNLNVEGGSSENN---2146
Qy 1061 FGSIIESPSTYFQEDADLDQGDQDDSRQNNSLNQEAGLLKQKLAILLGNQFTQYYQ 1120
Db 2147 -HSIKVSPYLS-QFQOD-----KQQLVLGTVKSVLHENTHVLGRQ 2184
Qy 1121 QNDKEIEFELINVEKVSSELSFRVEFKLAKT-----LEDNGKT-IR 1159
Db 2185 ASPKVKWEIKGTETTSDDVPVKNIIEVCSTYSKDSENYFTEAVEATAKAFMEDDELTDK 2244

Qy 1160 VLSDETMSL-----IVNTTIEK---TPEMSAVPEVFDTKWVEQVDPRPTPLAAKTK 1206
Db 2245 LPSHATHSLFTCPENEBEVLNSRIGKRGRGEPILLVGEPSIKRNLNNEFDRIENQEKS- 2303
Qy 1207 FVLKFKDQIPVDGSGNISDKWLASIPLVTHQOMLRLSPVU---KTIRELGLKTEQOQQO 1262
Db 2304 --LKASKSPF--DGTIKDR---RLFMHH--VSLEPTICVPFRITKE-----2340
Qy 1263 QOQOQOQOQKAVRKEEBELETYNPKDEFNILNPLTKAH---RLTL---SNL-VNNDPN 1314
Db 2341 --ROETQNFNFA-----PQOEF-----LSKSHLYEHLTLEKSSNLAVSGHPF 2382
Qy 1315 YKIEDLKVIKNBAGDHLAFSLRANNIKRLMNTPTTFADYNPPFFYNNEDWRSIDKYLNN- 1373
Db 2383 YQVS-----ATRNEKMRHLITTRPTKVFVPPFKTSHFHRVEQCVRNI 2426
Qy 1374 -----KGNVSHQOQAAGNGQSGLIQRLNKNIKPE---TFTPA-----LIALKOR 1416
Db 2427 NLEENRQKQNDHGSDDSKKNKINDNEIHOFNKNNSNQAADVTTTKCEBEPLDLI-----2481
Qy 1417 NNTNLSNYS-D-KIIMIKPK-----YLVERSIGVPWSTGLDGVIGSEQTKDGTG 1463
Db 2482 --TSLQNRARDIQMRILKKQORQVFPQPSGLYLAKTST-LP-RISLKAAGVGPACSH 2537
Qy 1464 SSSQKGFDDFIQALGLKQVTE-----YHGK-----LGLSIRIPDPGNEIAKIKDASN 1511
Db 2538 KQLYTYGVSKHCIK-INSKNABSFQPHTEBYFKESLWTKGILADGG---WLIPSDNG 2593
Qy 1512 KGEELKLSYDLFKNYLNEYEKSPK-TAKGWTNHPQKEVPNPQKL-----PEN 1563
Db 2594 KAKGEB-----FYALCDTPGVDPKLSIRIWWYNH-----YRWIWLKLAAMECAPPKE 2641
Qy 1564 YLNLVLNPKWVTLNYSDFITNLFVEPEGSRGSGTKLKQVIQKOVNNNYADWGSAYLT 1623
Db 2642 PANRCLS-PERVLQKRYDTEI-----DRSRSAIKIMER-----DDTAATKL 2686
Qy 1624 FWYDKNIITNPQNVTANITADVFIKDVKBELBNTKLIAPNITQMW 1668
Db 2687 VLCVSDIIS-----LSANISSETSSNKTSS-ADTKVAILIETLDGW 2725

RESULT 16

US-09-044-908-2
; Sequence 2, Application US/09044908
; Patent No. 6124104
; GENERAL INFORMATION:
; APPLICANT: Tavtigan, Sean V.
; APPLICANT: Kamb, Alexander
; APPLICANT: Simard, Jacques
; APPLICANT: Couch, Fergus
; APPLICANT: Rommens, Johanna
; APPLICANT: Weber, Barbara
; TITLE OF INVENTION: Chromosome 13-Linked Breast Cancer
; TITLE OF INVENTION: Susceptibility Gene
; NUMBER OF SEQUENCES: 124
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Venable, Baetjer, Howard & Civiletti
; STREET: 1201 New York Avenue N.W., Suite 1001
; CITY: Washington
; STATE: DC
; COUNTRY: USA
; ZIP: 22204
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/044,908
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:

QY	ASPDKPYMKDVQVQRYYLAKSQIOELLIKAKDYTKLAKLLSNRHTYINISLRLEQQLFDVN	632
1728	---DKNHL---SEKQDTYLSNS	1760
633	PRIPSSRDIEKAKFVLDKTEKNKYWQIYSSASPVFQNKMSLFGYYRYLLGLQPKQTIHEL	692
1761	GYLSKNKLDGIEBPVLKNVEDQK---NTSFSKVISNVKDANAY---PQTVNE-1806	
693	VKLQKAGLQPEGYENILPSDFNLDELKNIHRIKTPFLSQKDNFKLSLLDFNNYYDGIKAP	752
1807	---DICVEEL---VTSSSPCKNKAAIKLSISNSNF---BVGPP-1842	
753	EFGLPLFLPKELRRNSNSGGSQNSNSPWEQELISQPKD-----QNLNSQDLA	801
1843	AFRI-----ASKIVCVS---HETIKVKVDIPTDSFSKVIKNNENKSKIC	1885
802	QFSTKIW-----EKIIGDENEFDONNRLOQYKLLKDLQBSWINKTRDNLNYTYLGD	851
1886	Q---TKIMAGCYEALDDSEDIHNSLNDDECSHSHKVFADIQSEELIQHNQNM---SGLEK	1941
852	KLKVKPKN-NLEAKPRQISNLQELLTAFTYSAALSNNNNYYQDSGAKST-----899	
1942	VKISPCDVSLETSICKCSIGKL---HKSVSANTCGIFSTAGSKSVQVSDASLQNA	1997
900	IIFBEIABDPKVKVKY-----GADYVQLKFHYAIGFDP---NACKFQOEVISSRTI	950
1998	QVFSIEDSTKQVFSKVLFPKSNEHSDQLTRENTAIRTPEHLISQKSGFSYNNVNSAFSG	2057
951	YLKTSKSKLEADTIDQLNQAVKNAPLGLOSF-----YLDTERFCVQFKLATS	1000
2058	FSTAGK---QVSIKESLHKVGV---LEEFDLIRTEHSLHVSPTSQNV-----2102	
1001	VOHKOKETPLKCLNNDGYTLIHDKLKKPVIPOISSPEKDNFEGKLNQNGSQNVNST	1060
2103	-----SKILPRVDRKRNPEHCNVSEMEKCTCKEFLKS-----NNLNVGSGSEN---	2146
1061	FGSIIESYFTSNFOEADLDDQODSDROGNSLNDQEBAGLKKQKLAILLGNQFYQYQ	1120
2147	HSIKVSPYLS-QFQD-----KOQLVLGTKVSLENTHVLGKEQ	2184
1121	QNDKEIBEIILNVEKVSLSFRVEFLAKT-----LEDNGKT-IR	1159
2185	ASPKNVKMEIGKTETFSDPVKTWIEVCSTYKDSYENYFTEAVEITAKAFMEDELTD	2244
1160	VLSDETMSL-----IVNTTIEK---TPMSAVPEVFDTKWVEQYDPTPLAAKTK	1206
2245	LPSHATHSLFTCPENEEVLSNSRIGKRGEPLILVGEPSIKRNLNNEPDRIIENQES-	2303
1207	FVLKPKDQIPVDGSGNISDKWLASIPLVITHQMLRLSPVV-----KTIRELGLKTEQ	1262
2304	--LKASKSTP---DGTIKDR---RLPMH--VSLRPTICVFPRTKE-----2340	
1263	QOQOQOQOQPKAVKAREELETYNPKDEFNLNPLTKAH---RLTL-----SNL-V	1314
2341	--RQBIQNPFTA-----PGQEF-----LSKSHLYEHLTEKSSSNLAVSGHP	2382
1315	YKIEDLKVIKNEAGDHQALAFSLRANNIKRLMNTPIITPADYPPFYNNEDWRSIDK	1373
2383	YQVS-----ATREKMRHLITTCRPTKVPFPVPPKTSKSHHVEQCVRN	2426
1374	-----KGNVSSHQQAAGNQGSGLIQRLNKNKPE---TFTPA-----LIALKDR	1416
2427	NLEENRQONIDGHGSDSKNKINDNEIHOQNKNSQAAAVTPTKCEEPFLDI-----2481	
1417	NNTNLSNYSVD-KIMIKPK-----YLVERSIGVPWSTGLDGYISSEOTKDGT	1463
2482	--TSLQARDIQDMRIKCKQRQRPVPPQPGSLYLAKTST-LP-RISLKAAGVQVPSAC	2537
1464	SSSQQKGFDDPIQALGKNTK-----YHGK-----LGLSIRIFQGNELAKIKDAS	1511
2538	KOLYTVGVSKHCIK-INSKNAESFQFHTEDYFGKESLWTKGQLADGG---WLI	2593

Db 1758 SDSKFRKYSQKKAQAEQAKIOQKVNMIQKIKTNKALNSAQAQALBELKQAKL 1817
Qy 1493 SIRIFDPCNEL-----AKIKDASNK-----KGEKLLKSYDL 1524
Db 1818 -ISVQDVRELQKLVQSKVDTELKSTIEKSSSTQGIKQVDNKSMTEDDEKVKYYSK 1876
Qy 1525 PKNYLNEYKSPKPIAGKWTNHPDQKBYPNPNQKLPENYLNVLNQPWK-----VTL 1577
Db 1877 QIKLIQOQKQKAKYIKQLEBQKKAAGFPDIOEQITEWQN-----WKDQKQDFNLEL 1930
Qy 1578 YNSDDFTNLN-----VEPEGSDRGSGTKLKQV----- 1605
Db 1931 YNTKSGINDIYKSLADEVUSYKEMEKMRDIELEAHQKATQDLIDEIDKTDDEAKPQKE 1990
Qy 1606 -----IQK-----QVNNYADMGSAYLTFWYDKNIITNQPNVITANI-----ADVFID 1649
Db 1991 LKERQDSIQKLTQDQINQYSLD-----DSEFGSKVKVELTEQKQLEQLDLDLFDK 2040
Qy 1650 -----VKEEDNTKLIAPNITQWPNISGSKEF-----YKPTVFFGNW 1688
Db 2041 RESNKRKEALQDLEKDEESINNKYDNLVNDERAFKLEDKIMNGKITDIKQALNEFSKF 2100
Qy 1689 ENEN-SSMNSQAQPTTWEKIREGFALQAKSSFDQKTRTFVLTNAPLWKYGLGFQ 1747
Db 2101 INTNMESIGKISNNLIDKLE--ASNALNTAVKGNITGKKVSSFAS---GGYGTGLGA 2155
Qy 1748 GPNFKTDWRVLVFQND-----NOTAALRVQEQDRPEKSSBDKQKQW---IKPKVWIPE 1799
Db 2156 GKLAFLHDKELINKNTDTANILDTVKA VRETAVD-----DSPKWQGVKLADLIKK 2206
Qy 1800 EMFNSGNTRFVGMQIQGPNLTMLPVINSVVIYDFYRGTGSDNDVANLNAVQVKTIAP 1859
Db 2207 -----GITSIPS-----LVENNQSNL-----TNSLIPNLKIEIPSKTIAS 2243
Qy 1860 TNNAFNVPKERNISKI 1877
Db 2244 SGDKTINLTNTHIDKLI 2261
RESULT 18
US-09-932-183A-2
; Sequence 2, Application US/09932183A
; Patent No. 6833265
; GENERAL INFORMATION:
; APPLICANT: Estell, David A.
; TITLE OF INVENTION: Proteases From Gram-Positive Organisms
; FILE REFERENCE: GC394C1-US
; CURRENT APPLICATION NUMBER: US/09/932,183A
; PRIORITY FILING DATE: 2001-08-17
; PRIOR APPLICATION NUMBER: US 09/308,375
; PRIOR FILING DATE: 1999-05-14
; PRIOR APPLICATION NUMBER: PCT/US98/18828
; PRIOR FILING DATE: 1998-09-08
; PRIOR APPLICATION NUMBER: EP9719636.4
; PRIOR FILING DATE: 1997-09-15
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 2285
; TYPE: PRT
; ORGANISM: Bacillus subtilis
US-09-932-183A-2
Query Match 2.6%; Score 252; DB 4; Length 2285;
Best Local Similarity 18.1%; Pred. No. 5.5e-08;
Matches 395; Conservative 339; Mismatches 838; Indels 606; Gaps 94;
Qy 60 ANFTSDYQSVKALLNGKTFDPKSEFTDFVS---KPDFLTNNGRTVLEIPK-KYQVVIS 115
Db 330 AELTQTFKSMSTYLSGSLFYGAISGLKEMVQAIEIDTMTNIRRVNNEPDYKYNELL- 388
Qy 116 EFSPEDDKFRGLGHLEKLEDE-----GNIAQSATKFIYLLPLDMPKALGQYSYVDK 170

Db 389 -----QESIDLGDTLNKNITDILQMTGDFGRMGFDESELSTLTAKTAQVLQNVSLTPD 441
Qy 171 NFNNLIHPLNSFSAOSIKPLALTRSSDFIAKLNQFNQDELWVYLEKFPDL--EALKAN 228
Db 442 DTVNTLTAAMLNFN-----IAANDSISADKLNVEYDNN-----YAVTTLDLANSIRKAG 490
Qy 229 IRLQT-----ADFSFEKGNLVD--PFVYSFIRPNQONQKQEWASDLNQDQK 270
Db 491 STASTFGVELNDLIGYTTALTATTTRESGNI VGNLSKTI FARIGNQSSIKALSGIGSVK 550
Qy 271 TVRLYLRTFSPQAKTILKDYKYKDETFLLSSIDLKASNGTS-----LPANENDLKQDLVDL 327
Db 551 TAGGEAKS-----ASDLISEVAGKWDTLSDAQONTSIGVAGIYQLSRFNNMNFISI-- 603
Qy 328 LDVSDVFGGSEITNSQVQKVPASERSLKDRAVKFKDOOKPRIEFSLVEYDALSFYS 387
Db 604 -----AQNAAKTAANSTGSAMSEQQKYADSLQARVKNLQNNFTFEFAAASDA--FIS 653
Qy 388 QLQELVSKPNSIKDLVNATLARNLRFSLGKYNELFDDLASHLDYFVFLVSKAKIQSSITK 447
Db 654 --DGLHFTQAAGSLNASTG-----VIKSVGFLPPLAA-----VSTATLLLSKNTR 699
Qy 448 KLFIELPIKISLSKSIILGDQEPNI KTL-----FKEVTFKLDFRDVIEIKAF-----GLL 498
Db 700 TL-----ASSLIILGTRAMGOETLTATAGLEAGMT-----RAAVASRVLTALRGLL 744
Qy 499 YP-----GVNEELBOARQAQASFEKSKKGLKFEFSQOKEENSKAINNQEGLEDDNITE 554
Db 745 VSTLVGGAFAALGWALESLSF--AEAKKAKDDFEQSQQTNVBEATTNK-----DSTDK 797
Qy 555 RLPENSIQYQOENAGIGASPDKPY-----MIKQVQNYLYLAKSOIELIK 601
Db 798 LIQYKELQVKKRSRSLTSDBEQEYLQVTOQLACTFPALVKGYSQGNAILKTN-KELEK 856
Qy 602 AKDYTKLAKLLSNRHTYNI SLRLKEQLFDVNPRIIPSSRD-IEKAFVLDTKRNKYWOIY 660
Db 857 AIENTKEYLAKKQETRDSA---KKTPEDAKEIKKSKBELKQYKQIADYNDKGR----- 908
Qy 661 SSASPVFNKWSLFG-YRYLLGLD-PKQIHELKVLGQKAGLQFEGYENLPSPDNLEDL 718
Db 909 -----PKWDLIADDDDYKVAADKAKQSM-----LKAQSDIES- 940
Qy 719 KNIRIKTPLFSQKDNFKLSLLDFNNYDGEIKAPEGLPL---FLPKELRRNSNSGSGQ 775
Db 941 GNAVKDSVLISANAY--SSIDISNTLKTSLSDVVNKLNLKDDLDPELEKFSLSGLKQ 998
Qy 776 NSNSPWEQEIISQPKDQNLNSQDQLAQFSTKIWEKIIIGDENEPDQNNRLQYLLKDOES 835
Db 999 EK-----NQKALDSGDEKAFDNAKKDLQSLLETYSKSDSSIDVFKMSFDKAKNIKDGDKS 1054
Qy 836 WI-----NKTRNLTYWTVLGDYK--VKPNKLEAKERQIS----- 869
Db 1055 LSSVKSEVGDIGETLABAGNEAED-----FGKKLKEALDANSVDDIKAAIKEMSDAMQF 1108
Qy 870 -NLOELLTAFYTSAAALSNNWYODSGAKSTIIFEEIAB-----LDPKVKKEVG 917
Db 1109 DSVQDVLN-----GDFNNTKQDQVAPLNDLLEKWAEGKSIANEANTLIQKDE--- 1157
Qy 918 ADVYQLAKFHYAIGFDDNAGFNQ-EVIRS-----SSRTIYLTSGSK-LE 961
Db 1158 -----LAQAISIENGWVKINRDEVIKQKRVKLDAYNDVMYVSNKLMKTEVNNAIKTLN 1210
Qy 962 ADT--IDOLNOAVKNAPLGLOSFYLDTERFGVFOKLATSLAVQHOKKEKTLPKLNNQGY 1019
Db 1211 ADTLRIDSLKRLKERKLDMS-----EAELSDLVKSINNADAKKELKKEKMLQPGY 1266
Qy 1020 TLIH-----DKLKKPVIPOISSPEKDWFEGLKNGQSGQNVNVSTFGSIIIESPFSTNFQ 1075
Db 1267 SNSQIEMQSVKSALESYISASEAT-----STQENWKO---ALVEA---GTSLE 1310
Qy 1076 EDADLDQGDQDSRQGNNSLDNQBAGLQKLAILLGNQFTQYQOQNDKEIEFEIINVEK 1135

[illegible]

RESULT 19

RESULTS 15
US-08-480-604A-6

```

; Sequence 6, Application US/08480604A
; Patent No. 5736139
; GENERAL INFORMATION:
; APPLICANT: KINK, JOHN A.
; APPLICANT: THALLEY, BRUCE S.
; APPLICANT: PADHYE, NISHA V.
; APPLICANT: FIRCA, JOSEPH R.
; APPLICANT: STAFFORD, DOUGLAS C.
; TITLE OF INVENTION: VACCINE AND ANTITOXIN FOR TREATMENT AND
; PREVENTION OF C. DIFFICILE DISEASE
; NUMBER OF SEQUENCES: 32
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MEDLEN & CARROLL, LLP
; STREET: 220 MONTGOMERY STREET, SUITE 2200
; CITY: SAN FRANCISCO
; STATE: CALIFORNIA
; COUNTRY: UNITED STATES OF AMERICA
; ZIP: 94104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/480,604A
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/422,711
; FILING DATE: 14-APR-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/405,496
; FILING DATE: 16-MAR-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/329,154
; FILING DATE: 25-OCT-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/161,907
; FILING DATE: 02-DEC-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/985,321
; FILING DATE: 04-DEC-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/429,791
; FILING DATE: 31-OCT-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: INGOLTA, DIANE E.
; REGISTRATION NUMBER: 40,027
; REFERENCE/DOCKET NUMBER: OPDH-01763
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 705-8410
; TELEFAX: (415) 397-8338
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2710 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-480-604A-6

```

09	272	VRLYRTEFS---POAKTILKDY---KY-KDETFLSS-----IDLKASNG 309
----	-----	---


```
Db 185 FINTYKQINQKPTVPTDDIIKSHLVSEYARDETIVLESYRTNSLRKINSHGDIRAN-- 242
Qy 310 TSLFA-----NENDLKQDLVDLLDVSDFYGGQSETITTSNQVQKVPASERS 356
Db 243 -SLFTEQELLNIYQELLNRGNLAASDIVRLALKNFQ----- 281
Qy 357 LKORVKPKQOQKRIEKFSLY-BYDAL-SFYSOLOELVSKPNSI----- 399
Db 282 -----VYLDVMLPGIHSOLFRTIHRPSSIGLORWEMIKLEAIMK 321
Qy 400 -KOLVNATLARNRFLSLGKYNFLPDDLASHL-DYFVLVSRAKIKOSSITKKLFIELPIKI 457
Db 322 YKKTINNTSEN-----FDKLDQOLKDNFKLIIESKSEKSIKSKL----- 362
Qy 458 SLKSSILGDQBPNIKTLFEKEVTFKLNFRNDRVEIEKAPGLYPGVNBELOARQAQASPF 517
Db 363 -----ENL-----NVSDEIKIAFAL-----GSVINQALISKQGSY 393
Qy 518 EKEKSKGLKEFSQOKEBNSKAINNOE---GLEDDNITE-----RLPENSPI 562
Db 394 -----LTLVLVIEQVKNRYQFLNQHLPNPAIESDNNFTDTTKIFHDSLFNSATAENSMF 445
Qy 563 -----QYQOENAGLGASDKPYMIK---DVONQRY---LAKSOIQELIKADYTKLA 609
Db 446 LTKIAPYLQ-----VGFMEARSTISLSPGAYASAYDFINLQENTIEKTLKASD----- 496
Qy 610 KLLSNRHTYINSLRLKEQLFDPNPRIPSSRDIEKAFVLDKTEKNKYWQIYSSASPVQON 669
Db 497 -LIEFKPPENNLSQLTEQ--EINSL--WSPDQASAKYQFEKYVRDYTGGSLSBNGVDNF 551
Qy 670 KWSLFGYRYLLG-----LDPKQTIHELKLGKAGLOPEGYENLPSDNLEDLK 719
Db 552 KNTALD-KNYLLNNKIPSNVVEEAGSKNYHYIIQL-QGDDISYEATCNL---FSKNPKN 606
Qy 720 NIKIKTLPFSQDNFKL-----SLLDNNYDGEIKAPE-----FGLPLFL 760
Db 607 SIIOQNNVEAKSVFLSDGESILENKY-----RIPERLKNKEKVKVTFIGHGKDEFN 661
Qy 761 PKELRNSNNGGQSN-----SPWQEI-----ISOFKQONLSNQDLAQFSTKI 807
Db 662 TSEFARLSVDSLNEISSFLDTIKLDISPKNVEVNLGNCNMFSDNVEETYPGKLLLSI 721
Qy 808 WEKIIGDNEFDQNN-----RLQYKLLKOL---QESWINK-----TRNLY 845
Db 722 MDKITSTLFDVKNKSIITIGANQYEVRIKNSGKELLASHGKWINKEBEAIMSDLSKEYIF 781
Qy 846 WTYLGDKLKVKPKNNLEAKFROISNLQELLTAFYTSALS-----NN----- 887
Db 782 FDSIDNKLKAKSKN-----IPGLASISGDIKTLILDASVPDTKFLNNLKNLTISSIGD 836
Qy 888 WNYQDSGAKSTIIFEEIAEL--DPKVKKEKGVADYVOLK-----FHYAIFDD---NAG 936
Db 837 YIYKEKLEPVKNIHNSIDDLIDEFNLENVSDLEYELKLNLDKYLISFEDISKNS 896
Qy 937 KPNQEVIRSSRTIYLTSGK-----SKLEADRID 966
Db 897 TYSVRPINKNGESVYVETEKEIFSKYSEHITKEISTIKNSIITDVGNGNLLNDIQDHTS 956
Qy 967 QLNQAVKQAPLQSFYLDTRFQFKLATS LAVQ----- 1002
Db 957 QVN--TLNAAFFIQSLIDYSSNKKVDNLNSTSVKQVLAQLFSTGINTIYDSIQLVNLS 1014
Qy 1003 -----HKQKEKT----- 1009
Db 1015 NAVNDTINVLPITTEGIPIVSTILDINGLGAAILKELLDEHDPLLKKELEAKGVLAINMS 1074
Qy 1010 -----LPKLNNDGYTLIHKLKKPV---IPQISS 1037
Db 1075 LSIATVASIVGIGAEVITFLPIAGISAGISPLVNE--LTLHDKATSVVNYFNHLS 1132
Qy 1038 -----PEKDWPEKLNQ-----GOSQNVNVTFGS- 1063
```

```
1133 KKYGLPKTDDDKILVPIDDLVISEIDFNNNNSIKLGTCLNLAAMEGSGHVTVTGNIDHFFSS 1192
1064 -IIESYFSTNFQEDADLDODGODDSRO-----GNSLDNQEAGL 1102
1193 PSISSHIPSLIYSAIGIETENLDFPSKIMMLPNAPSRVFWETGAVPGLRSLENDGTRL 1252
1103 LKOKLAILLGNQPIQYQQNDKEIE-----FEIINVEKVSLSFR-----VEFK 1146
1253 LDSIRDLPGKFVRFVAFPDYAITTLKPVYEDTNIKIKLDKDRNPFIMPTITITNEIRNK 1312
1147 LAKTLEDNGKTIKRVL-----SDETMSL-IWNTTIEKTPMSAV 1183
1313 LSYSPFCAGGTYSLLSYSPISNTINLSKODLWIFNIDNEVREISIENTGIKKGLKIDV 1372
1184 PEVDTKWVEQYDPRTPLAAKTFVLKFQDQIPVDGSGNISDK-----WL 1228
1373 LSKIDIN-----KNKLI---GNQIDFSGDIDNKDORYIFLTCELDKISLI 1416
1229 ASIPLVIHQOML-----RLSPVVKVTIRELGLK-----TEQOQOQOQOQOQOQ 1269
1417 IEINLVAKSYSLLSGDKNYLISNLSNTIEKINTLGLDSKNIAVNYTDESNNKYFGAISK 1476
1270 QPOKXAVRKEE---LETYNP-----KDEFNILN----- 1295
1477 TSOKSIIHYKKDSKNILEFYNDSTLEFNSKDFIAEDINVFMKDDINTITGKYVYDNNTDK 1536
1296 -----PLTKAHLTLNLSNVNDPNYKIEDLVKNEAGDHLAFSLRANNIKRLMNTPI 1349
1537 SIDFSISLVSKNQVKNGLYLSNESVYS-SYLDPFVKNSDGH-----NTSNFMNLFIDNI 1589
1350 TFADYNPFYNNEDWRSIDKYL-----NNKG-----NVSSHQOQAA 1385
1590 SF--WKLFGFENINF-VIDKYFTLVGNTLGYVEFICDNNKNIDIVFGEWKTSSKSTIF 1646
1386 GGNQGS-----GLIORLKN-NIKPETFTFALIALKDRN 1417
1647 SGGRNVVVEPIYNPDTGEDISTSLDFSYPGLYIDRYINKVLAPDLYT-SLINI----- 1701
1418 NTNL--SNYSKIIIMKPK-----YLVERSIGVPWSTGLDGYIGSEQTKDGTSSSSQ 1467
1702 NNTYSSNEYPEIILVNPNTFHKKNINLDBSSFEYKWSSTEGSDFILVRYLESNNKILQ 1761
1468 Q---KGPDDQFIQALGLKNTYHKGKGL-----SIRIFDPGNELAK----- 1505
1762 KIRIKGI-----LSNTQSFNKSIDFKD IKKLSLGYIMSFKNSFENSELDRHLG 1812
1506 IKDASNK---KGEKLLKSYDLFPQVLYNEVEKSPKIAKGTWNIHDPQKEYPNPKLP 1561
1813 FKIIDNKTYVYDEDSKLVKGLININNSLFPDPIEFNLVTGWQTI--NGKKY----- 1862
1562 ENYLNVLVNQPKVTLX---NSSDFITN-----LFVBEPSGDSRSGTKLKQVIOKQV 1610
1863 --YFD--INTGAULTSKYIINGKHIFYNNDBVMOLGVFKGPDGFEYFAPANTQ----- 1911
1611 NNNYAD---WGSAYLT-----FWYDKN-----IITNQPNVITANIADVFIKDKVEL 1653
1912 NNNIEGQAVVQSKFLTLNGKKYFDDNNSKAVTGWRIINNEKYVFNPNNA-IAAVGLQVI 1970
1654 EDNTKLIAPN---ITQWPNISGSKKPYKPTVFFGN 1687
1971 DNKKYFNPDPAIISKGMQTVNGSRYVFTDITDAIAPN 2007
```

RESULT 20
US-08-405-496A-6
; Sequence 6, Application US/08405496A
; Patent No. 5919665
; GENERAL INFORMATION:
; APPLICANT: WILLIAMS, JAMES A.
; TITLE OF INVENTION: VACCINE FOR CLOSTRIDIUM BOTULINUM
; TITLE OF INVENTION: NEUROTOXIN
; NUMBER OF SEQUENCES: 30
; CORRESPONDENCE ADDRESS:

Qy 1229 ASIPLVTHQML-----RLSPVKTIRELGLK-----TEQOQQOQQOQQOQ 1269
Db 1417 IEINLVAKSYLLSGDKNYLISNLSTIEKINTGLDLSKNIAYNTDESNNKYFGAISK 1476
Qy 1270 QPQKAVRKEE---LETNP-----KDFNILN----- 1295
Db 1477 TSQSIHYKDSKNILEFYNDSTLEFNSKDFIAEDINVPKODINTITGKYVYDNNTK 1536
Qy 1296 -----PLTKAHLITLNLVNDPNYKIEDUKVTKNAGDHQHLAFSRANNIKELMTPI 1349
Db 1537 SIDSISLVSKNQVKNGLYLNESVYS--SYLDFVKNSDGH-----NTSNFNLFLDNI 1589
Qy 1350 TFADYNPFYNYEDWRSIDKYL-----NNKG-----NVSSHQOQAA 1385
Db 1590 SF--WKUFGFENINF-VIDKTYFTLVGKTNLGYVFIQNNKNIIDYGEWKTSSKSTIF 1646
Qy 1386 GGNQGS-----GLIQRLNK-NIKPFTFPALIALKDRN 1417
Db 1647 SGNRNVVVEPIYNPDGTGEDISTSDFSYELYGIDRVINKVLIAPDLYT-SLINI--- 1701
Qy 1418 NTNL--SNYSKILIMIKPK-----YLVRSIGVPMWSTGLDGYIGSEQTKDGTSSSQ 1467
Db 1702 NTVYSNEYYPEIIVLNPTFHKKVNINLSDSSPEYKMWSTEGSDFILVRYLEESKKILQ 1761
Qy 1468 Q---KGFQDQFIQALGLKNTYHKGKGL-----SIRIFDPCGNELAK--- 1505
Db 1762 KIRIKGI-----LSNTQSFNKSIDFKIKKLSGYTMSNFKSFNSENEIDRHLG 1812
Qy 1506 IKDASNK---KGBEKLKSYDLFQVLYNEYEKKSPIAKGWTNIHPDQEPNPNQKLP 1561
Db 1813 FKIDNKTYDEDSKLVKGLININNSLYFDPFIEFLVGTQWII--NGKXY----- 1862
Qy 1562 ENYMLVLNPKWVTLY---NSSDPITN-----LFVEPEGSDRGSGTKLKOVIQKV 1610
Db 1863 --YFD--INTGAALTSYKIINGKHFEYFNNDGVMQGVKFGPDGFEYFAPANTQ----- 1911
Qy 1611 NNNYAD---WGSALVT---FWYDKN-----IITNQPNVITANIADVFVKVKEL 1653
Db 1912 NNNTEGQAIYQSKFELTNGKKYFDDNNSKAVTGWRIINNSKYFNPNNNA-IAAVGLQVI 1970
Qy 1654 EDNTEKLIAPN---ITQWPNISGSKEFEYKPTVPFGN 1687
Db 1971 DNKKYFNPDTALISKGHTVNGSRYYEDTDTAIAFN 2007

RESULT 21

US-08-915-136-6

Sequence 6, Application US/08915136

Patent No. 6290960

GENERAL INFORMATION:

APPLICANT: KINK, JOHN A.
APPLICANT: THALLEY, BRUCE S.
APPLICANT: PADHYE, NISHA V.
APPLICANT: FIRCA, JOSEPH R.
APPLICANT: STAFFORD, DOUGLAS C.
TITLE OF INVENTION: VACCINE AND ANTITOXIN FOR TREATMENT AND PREVENTION OF C. DIFFICILE DISEASE
NUMBER OF SEQUENCES: 32
CORRESPONDENCE ADDRESS:
ADDRESS: MEDLEN & CARROLL, LLP
STREET: 220 MONTGOMERY STREET, SUITE 2200
CITY: SAN FRANCISCO
STATE: CALIFORNIA
COUNTRY: UNITED STATES OF AMERICA
ZIP: 94104

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/915,136
FILING DATE:

CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/480,604
FILING DATE:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/405,496
FILING DATE: 16-MAR-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/329,154
FILING DATE: 25-OCT-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/161,907
FILING DATE: 02-DEC-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/985,321
FILING DATE: 04-DEC-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/429,791
FILING DATE: 31-OCT-1989
ATTORNEY/AGENT INFORMATION:
NAME: INGOLIA, DIANE E.
REGISTRATION NUMBER: 40,027
REFERENCE/DOCKET NUMBER: OPHD-01763
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 705-8410
TELEFAX: (415) 397-8338
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 2710 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-915-136-6

Query Match 2.6%; Score 250.5; DB 3; Length 2710;

Best Local Similarity 18.1%; Pred. No. 8.9e-08;

Matches 387; Conservative 274; Mismatches 647; Indels 829; Gaps 101;

Qy 180 LSNFSAQSIKPLATRSDF--TAKLNQFNQDLSLW-----YLEKFFDLEAKA 227
Db 71 LSNLKDILKEVILIKNSNTSPVEK-----NLHFVWIGGEVSDIALEYIKOWADINA-EY 124
Qy 228 NIRLOTADFSPKGNLVDPPFYSP-----IRNPQ--NOKEWASDLN---QDQKT 271
Db 125 NIKLWYDSEAFVNTLKAIVESSTTEALQLLEBEEIQNPQPDNNKFYKRMFEITYDQKR 184
Qy 272 VRLYLRTFS---PQAKTILKQY---KY-KDETFLSS-----IDLKASNG 309
Db 185 FINYKSGINKPTVPTTIDDIKSHLVSEYNRDETIVLESYRTNSLRKINSNHHGIDIRAN-- 242
Qy 310 TSLEA-----NENDLKDQDLDVLDVSDYFGQSETITTSNQVKVPASERS 356
Db 243 -SLTFQELLNYSQELLNRGNLAAASDI-VLLALKNFGG----- 281
Qy 357 LKDRVFKKQDQKPRIEKFSLY-EYDAL--SFYSQLOELVSKPNSI- 399
Db 282 -----VYLDVDMLPGHISDLFTKISRPSIGLDRWEMIKLEAIMK 321
Qy 400 -KDLVNATLARNRFSIGKYNFLDDELASHU--DYFVLVSKAKIKQSSITKGLFIELPIKI 457
Db 322 YKVINNYTSEN-----FDKLDQQLKDNFKLIIESKSEKSEIFSKL----- 362
Qy 458 LKSSILGDQEPNIKTLFEKEVTFKLDNFRDVELEKAFGLLYPGVNELEQAKQAQASAF 517
Db 363 -----ENL-----NVSDLEIKIAPAL-----GSVINQALTSKQGSY 393
Qy 518 EKEKSKGLKEFSQOKKENSAINNQE---GLEBDDNITE-----RLFPENSPI 562
Db 394 -----LTNLVIEQVKRYQFLNQHLNPAIESDNNFTDTTKIFHDSLFNSATAENSME 445
Qy 563 -----QYQENAGLGASPKPYMIK-----DVQNGRY--LAKSQIQLKAKDYTKLA 609
Db 446 LTKIAPYLQ-----VGFMEARSTISLSPGAYAYDYFINLQENTIEKTLKASD----- 496

QY 610 KLSNRHTYNI SURLKQLFDVNPRI PSSRDIKAKPVLDTKTEKNKYQIYSSASPVQON 669
DB 497 -LIEFPENNLSQLTEQ--EINSL--WSPQASAKYQEKYVRDYTGGSLSNDGVDEN 551
QY 670 KWSLFGYRVLG-----LDPKQTHLVLKQKAGLOFEGEYNLPSPDNLEDLK 719
DB 552 KNTALD-KNYLLNKNKIPSNVVEBAGSKVYHYIIQL-QGDDISYEATCNL---FSKNPKN 606
QY 720 NIKITPLFSQKONFKL-----SLLOFNYYDGEIKAPE-----FGLPLFL 760
DB 607 SIIQRMNESAKSYFLSDGSESILELKY-----RIPERLKNKEKVYTFIGHGKDEEN 661
QY 761 PKELRNSSNGSGQSN-----SPWEQBI-----ISQFKQNLNQDQLAQFSTKI 807
DB 662 TSEFARLSVDSLSEISFLDTIKLDSIPKQVENVLLGCMFSDYFNVEBETPGKLLSI 721
QY 808 WEKIIGDENEFDONN-----RLOYKLLKDL--QESWINK-----TRDNLX 845
DB 722 MDKITSLPDVNKNSTITIGANQYEVNRINSEGRKELLAHSGKWKINBEAIMSDLSKEYIF 781
QY 846 WYILGOKLKVKNLEAKPROISNLQELLTAFYTSAAALS-----NN-----887
DB 782 FDSIDNKLKAKSN-----IPGLASIEDIKTLILDASVSDTKFILNKLNIIESSIGD 836
QY 888 WNYQDSGAKSTIFBEIABL--DPKVKKVGADVYOLK-----PHYAIGFDD---NAG 936
DB 837 YIYKLEPKNIIHNSIDLDIDFNLENVSDLEYELKLNLDKYLISFEDISKNS 896
QY 937 KFNQEVIT-RSSRTIYKTSK-----SKLEADTID 966
DB 897 TYSVRFINSGESVYVETEKEIFSKYSEHITKEISTIKNSIITDVGNNLNDIQDHTS 956
QY 967 QLNQAVKNAPLQSGFYLDTERFGVFOKLATS LAVO-----1002
DB 957 QVN--TLNAAFFQSLIDISSNKDVLNDLSTSVKQVLAQFSTGLTYDSIQLVNLIS 1014
QY 1003 -----HKOKEKT-----1009
DB 1015 NAVNTINVLPTITEPIVSTILDGINLGAALKELLDEHDP LLLKKELEAKVGVLAINMS 1074
QY 1010 -----LPKLANGDVTLHDKLKKPV--IPOISS 1037
DB 1075 LSTAATVASIVIGAEVITFLPIAGISAGIPSLVNE--LILHDKATSVVNVFNHLS 1132
QY 1038 -----PEKOWFEGLKMN-----GQSONVNVSTFGS- 1063
DB 1133 KKYGLKTEDDKTLVPIDDLVISEIDFNNSIKLGTCLNLA MEGSGHVTGNIDHFFS 1192
QY 1064 -ITESPYFTNFQEDADLDQDQDSDRQ-----GNNSLDNQEAGL 1102
DB 1193 PSISSHIPSLSIYSAIGIETENLDFSKIMPLNAPS RVFWMETGAVPGLRSLNDGTRL 1252
QY 1103 LKQKLAILLGNQFIQYQNDKIE-----FEIINVEKYSLSFR-----VEFK 1146
DB 1253 LQSI RLYPGKFYWRFYAFDYAITTLKPVYEDTNIKI LDKDTRNFIMPTITTEIRNK 1312
QY 1147 LAKTLEDNGKTI RVL-----SDTMSL-IVNTTIEKTPEMSAV 1183
DB 1313 LSVSFDGAGGTYSLLSSYPIS TNNLSKDDLWIFNIDNEVREISIENTGKGLIKDV 1372
QY 1184 PEVFTKWVEQYDPRTPLAAKTFVLKFKDQIPVDSGNSDK-----WL 1228
DB 1373 LSKIDIN-----KNKLIIT--GNQITDFSGDIDNKDRYIFLTCELDKISLI 1416
QY 1229 ASITPLVHOOML-----RLSPVVKTI RELGLK-----TEQQOQOQOQOQ 1269
DB 1417 IEINLAKSYLLSGDKNVLNLSNTIEKINTLGLDSKNIA YNTDESNNKYFGAISK 1476
QY 1270 QPOKAVRKEE-----LETYNP-----KDEFNILN-----1295
DB 1477 TSQKSIHYKDKSKNILEFVNDSTLBFNSKDFTAEDINVMKDDINTITGKYVVDNNTDK 1536

QY 1296 -----PLTKAURLTSLNUNNDPNYKIEDLKVIKNEAGDHQALFSLRANNIKRLMNTPI 1349
DB 1537 SIDFSISLVSKNQVKNGLYNLESVYS--SYLDFVKNSDGH-----NTSNFMNLFLDNI 1589
QY 1350 TFADYNPRFYNNEDRSIDKYL-----NNKG-----NVSSHQOQAA 1385
DB 1590 SP--WKLFGFENINF-VIDKYFTLVGKTNLGVVEFCDDNNKIDIDYFGWKVSSSKSTIF 1646
QY 1386 GNGQGS-----GLIQR LNK-NIKPETTFPALIALKDRN 1417
DB 1647 SGNRNVVVEPIYNPDGTGEDISTSLDVSFVPLYGIDRYINKVLIAPDLVT-SLINI--- 1701
QY 1418 NTN L--SNVSDKIIMKPK-----YLVERSIGVPSWSTGLDVGIGSQTKDGTSSSQ 1467
DB 1702 NTNYSNEYPIEIIVLNPNTHKKNINLNDSSSFYKWNSTEGSDFLVRYLEESNKILQ 1761
QY 1468 Q--KGFDQDFI QALGLKNTVEHGKGL-----SIRIFPGNELAK--- 1505
DB 1762 KIRIKI-----LSNTQSFNMSIDFKDKKLSGLYIMSNF KSFNSENELDRHLG 1812
QY 1506 IKDASNK-----KGEKLLKSYDLFKNYLMEYKKS PKIAKGTNTHPDQOKEYPNPNQKLP 1561
DB 1813 FKIIDNKTYVYDEDSKLVKGLININSLFYDPDIEFENLVGTQTI--NGKYY----- 1862
QY 1562 ENYLNVLNQPWKVTLY--NSSDFITN-----LFVEPEGSDRGSTKUKQVIOKQV 1610
DB 1863 --YFD--INTGAALTSYKIIINGKHFYFNNDGVMLGVLFGKPDGFEYFAPANTQ----- 1911
QY 1611 NNNYAD--WGSAYLT-----FWYDKN-----IITQPNVITANTADVFKDKVKEL 1653
DB 1912 NNNISQALVYOSKFTLNGKKYIFDNNKSKAVTGHWRIINNEKYFNPNNNA-TAAVGLQVI 1970
QY 1654 EDNTKLIAPN---ITQWMPNISGSKEKYKPTVFFGN 1687
DB 1971 DNKKYFNPDTAII SKGQTVNGSRVYPTDTUTAIAPN 2007
RESULT 22
US-08-957-310-6
; Sequence 6, Application US/08957310
; Patent No. 6365158
; GENERAL INFORMATION:
; APPLICANT: Williams, James A.
; APPLICANT: Kink, John A.
; TITLE OF INVENTION: IDENTIFICATION OF NEUTRALIZING EPITOPES
; TITLE OF INVENTION: OF TOXIN A AND TOXIN B FOR THE TREATMENT OF C. DIFFICILE
; TITLE OF INVENTION: DISEASE
; NUMBER OF SEQUENCES: 22
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Medien & Carroll
; STREET: 220 Montgomery Street, Suite 2200
; CITY: San Francisco
; STATE: California
; COUNTRY: United States of America
; ZIP: 94104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/957,310
; FILING DATE: 23-OCT-1997
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/329,154
; FILING DATE: 24-OCT-1994
; APPLICATION NUMBER: US 08/161,907
; FILING DATE: 02-DEC-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/985,321
; FILING DATE: 04-DEC-1992
; PRIOR APPLICATION DATA:

```

; APPLICATION NUMBER: US 07/429,791
; FILING DATE: 31-OCT-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Ingolia, Diane E.
; REGISTRATION NUMBER: 40,027
; REFERENCE/DOCKET NUMBER: OPND-01121
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 705-8410
; TELEFAX: (415) 397-8338
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2710 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-957-310-6

Query Match      2.6%; Score 250.5; DB 3; Length 2710;
Best Local Similarity 18.1%; Pred. No. 8.9e-08;
Matches 387; Conservative 274; Mismatches 647; Indels 829; Gaps 101;

Qy 180 LSNFSAOSIKPLALTRSDSF--IAKLNQFNQDELWV-----YLEKFFDLEAKA 227
Db 71 LSNLKKDILKEVILKNSNTSPVEK-----NLHFVWIGGEVSDIALEYIKQWADINA-EY 124

Qy 228 NIRLOTADFSPEKGNLVDVPFVSP-----IRNPQ--NOKEWASDLN---QDQKT 271
Db 125 NIKLWYSEAFVNTLKKAIIVESSTTEALQLEEEIQNPQDNMKFYKGRMEFYDQKR 184

Qy 272 VRLVLRTEFS---POAKTILKDY---KY-KDETFLSS-----IDLKASNG 309
Db 185 FINYKQINQPTVPTDIDIIKSHLVSEYNDETVLESYRTSLRKNSNHGIDIRAN-- 242

Qy 310 TSLFA-----NENDLKQDLVDLLVDSDYFGGOSETITSNSQKVPVPSERS 356
Db 243 -SLTFEQELLNYSQELNLRGNLAASDIVALLAKNFG----- 281

Qy 357 LKDRVKFKDQOKPRIEKFSLY-EVDAL-SFYSQLQELVSKPNSI----- 399
Db 282 -----VYLDVMDLPGIHSDLFKTIISRPSSIGLDRWEMIKLEAIMK 321

Qy 400 -KDLVNTALNRLRSLGKYNFLPDLLASHL-DYVFLVSKAKIKOSSITKKLTLPKI 457
Db 322 YKKYINNTSEN-----FKLDQQLKDNFKLIIESKSEKSEIFSKL----- 362

Qy 458 SLKSSILGQEPNIKTLFEKEVTFKLDNFRDVEIEKAFGLLYPGVNEELEQAKAQRASF 517
Db 363 -----ENL-----NVSDLEIKIAFAL-----GSVINQALISKQGSY 393

Qy 518 EKEKSKGLKEFSQOKEKNSKAINQOE---GLEEDDNITE-----RLPENSPI 562
Db 394 -----LTNLVIEQVKNRYOFLNQHLPALPAIESDNFTDTTKI FHDLSLFNSATAENSMF 445

Qy 563 -----QYQOENAGLASDPKPMYK-----DVQNRYY---LAKSOIOELIKAKDYTKLA 609
Db 446 LTKIAPYLO-----VGFMEPEARSTISLSPGAVASAYYDFINLOENTIEKTLKASD----- 496

Qy 610 KLLSNRHTYINSLRLEQLDFVNPRIIPSRSDRIEKAFVLDKTEKNKYWQIVSSASPVFQN 669
Db 497 -LIEKFPENNLSQLTEQ--EINSL--WSFDQASAKYQFKEYVRYDTGGSLSEDNGVDFN 551

Qy 670 KWSLPGYVRYLIG-----LDPKQTIHLVLKGLKAGLOFPGYENLPDSDFNLEDLK 719
Db 552 KNTALD-KNYLLNNKIPSNVNBAGSKNYVHYIQL-QGDDDISYEATCNL-----FSKNPKN 606

Qy 720 NIKIKTPLPSQDNFKL-----SLLDNNYDGEIKAPE-----FGLPLPL 760
Db 607 SLIIQRMNESAKSYFLSDGESILELNKY-----RIPERLKNKEKVKVTFIGHKDEFN 661

Qy 761 PKELRRNSSNGSGQNSN-----SPWEQEI-----ISQPKQNLNQDLAQFSTKI 807
Db 662 TSEFARLSVDSLNEISSFLDTIKLDISPKNVEVNLGCMNMFSDFNVEETYPGKLLLSI 721

```

```

Qy 808 WEKIIGDENEFDONN-----RLQYKLLKDL---QBSWINK-----TRDNLV 845
Db 722 MDKITSTLPDVNKNSTIGANQVEVINSEGRKELLASHGKWINKEEAIMSDLSKEYIF 781

Qy 846 WTYLGDKLVKPKQNLKLEAKFRQISNLOELLTAPYTSAAALS-----NN----- 887
Db 782 FDSIDNKLKAKSKN-----IPGLASISEDIKTLLLDASVSPDTKFFILNNLKLIIESIGD 836

Qy 888 WNYQDSGAKSTIIPBEIAEL--DPKVKEKGVADVQOLK-----PHYAIGFOD---NAG 936
Db 837 YIYEKLEPVKNLIHNSIDDLIDEFNLENVSEDELKLNLDKYLISFEDISKNS 896

Qy 937 KPNQEVIRSSRTIYLTSGK-----SKLEADTID 966
Db 897 TYSVRPFINKSNGESVVVETEKEIFSVSEHITKEISTIKNSIITDVNGNLLDNIQLDHTS 956

Qy 967 QLNQAVKNAPLGLOSFYLDTEREGVQKLAISLAVQ----- 1002
Db 957 QVN--TINAAFFIQSLIDYSSNKDVLNLDLSTSVKQVLAQLFSTGLNTIYDSIQLVNLIS 1014

Qy 1003 -----HKQKEKT----- 1009

Db 1015 NAVNDTINVLPITTEGIPVSTILDGINLGAAIKELLDEHDPLLKLEBAKVGLAINMS 1074

Qy 1010 -----LPKKLNDGYTLIHDKLPV--IPOISS 1037
Db 1075 LSTAATVASIVGIGAEVTIFLLPIAGISAGIPSLVNN--LILHDKATSVVNYFNHLS 1132

Qy 1038 -----PEKOWFEKLNQ-----QSQNVNVSTFGS- 1063
Db 1133 KKYGPLKTEDKILVPIDDILVISEIDFNNSIKLGTNCLAMEGGSGHTVGTNIDHFSS 1192

Qy 1064 -ITESPYSTNFQEDADLDODGODDSRQ--GNNSLDNQEAGL 1102
Db 1193 PSISHTPSIUSIAGIETENLDFSKIMLWNPASRVFWETWGAVPGLRSLNDGTRL 1252

Qy 1103 LKQKLAILLGNQFTQYYQNDKEIE-----FEIINVEKVSLSFR-----VEFK 1146
Db 1253 LDSIRDLYPGKFYWRFAFDYAITLLKPVVEDTNIKIKLDKOTNFIMPITTTNEIRNK 1312

Qy 1147 LAKTLENGKTIKRVL-----SDETMSL-IUNTTIEKTPENSAY 1183
Db 1313 LSVSFDGAGGTYSILLSSYPISNTINLSKDBLWTFINIDNEVREISIENTGIKKGLKIDV 1372

Qy 1184 PEVEDTKWVEQYDRTPLAAKTFVLAKFKQIPVDGSGNISDK-----WL 1228
Db 1373 LSKIDIN-----KNKLI--GNQTFDSDIDNKDRIYFLTCELDKISLI 1416

Qy 1229 ASIPLVTHQOML-----RLSPVVKTIRELGLK-----TEQOQOQOQOQOQOQ 1269
Db 1417 IEINLVAKSYLLSGDKNYLISNLSNTIEKINTLGLDSKNIAANYTDESNNKYFGAISK 1476

Qy 1270 QPKKAVRKEBE--LETYNP-----KDEFNILN----- 1295
Db 1477 TSQSIHYKDKSNILEFYNDSTLEFNSKDFIAEDINVFMKDDINTITGKYVDNNTDK 1536

Qy 1296 -----PLTKAHLRTLNLNVNDPNYKIEDLKVINEAGDHOLAFSLRANNIKRLMNTPI 1349
Db 1537 SIDFISILSVKNQVKVNGLYLINESVYS--SYLDFVKNSDGHH-----NTSFMNLFLDNI 1589

Qy 1350 TFADYNPFYFYNEDWRSIDKYL-----NNKG-----NVSSHQOQAA 1385
Db 1590 SF--WKLGFENINF-VIDKYFTLVGLKTNLGVVFICDNNKNKNIIDYFGEWKTSKSTIF 1646

Qy 1386 GGNQGS-----GLIQLNK-NIKPEFTTALTALKORN 1417
Db 1647 SGNGRNVVVEPIYNPDTGSDISTSLDSFSEPLYGIDRYINKVLIAPDLYT-SLINI----- 1701

Qy 1418 NTNL--SNYSKIIIMKEK-----YLVERSIGVPWSTGLGVIGSEQTKDGTSSSQ 1467
Db 1702 NNTYVSNEYPIEIIVLNPNTHKKVNIINLSSSFYKWKSTEGSDFILVRYLEESNKILQ 1761

Qy 1468 Q----KGPDQDFTQALGLKXNTEYHGKGLG-----SIRIFDPGNELAK----- 1505

```

Db 1762 KIRIKI-----LSNTQSFNKMISDFKDKIKUSLGLVIMNFKSFENSELDRDLG 1812
Qy 1506 IKDASKN-----KGEEKLLKSYDLFKYVNLNEYEKSPKIAKGWTHIHDPDOKEYENPNQKLP 1561
Db 1813 FKIIDNKTYYYDEDSKLVKGLININNSLFYDFDPIEFNLVTGMQTI--NGKKY----- 1862
Qy 1562 ENYLNVLNPNWKVTLY---NSSDFTN-----LFVEPEGSDRGSGTKLKQVQKQV 1610
Db 1863 --YFD--INTGAALTYSKIINGKHFFVNDGVMLGVFKGPDGFEVFPAPANTQ----- 1911
Qy 1611 NNNYAD-----WGSAYLT-----FWYDKN-----IITNOPVITANTADVFKDKV 1653
Db 1912 NNNIEGQAIYQSKFLTLNKKYFYDNNNSKAVTGWRILINNEKYFNPNNNA-IAAVGLQVI 1970
Qy 1654 EDNTKLIAPN---ITOWMPNISGKFKYKPTVFFGN 1687
Db 1971 DNKKYFNPDTAISGMQTVNGSRYYFDDTAIAFN 2007

RESULT 23
US-10-011-366-6
; Sequence 6, Application US/10011366
; Patent No. 6573003
; GENERAL INFORMATION:
; APPLICANT: Williams, James A.
; Kink, John A.
; TITLE OF INVENTION: IDENTIFICATION OF NEUTRALIZING EPITOPES
; OF TOXIN A AND TOXIN B FOR THE TREATMENT OF C. DIFFICILE
; DISEASE
; NUMBER OF SEQUENCES: 22
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Medlen & Carroll
; STREET: 220 Montgomery Street, Suite 2200
; CITY: San Francisco
; STATE: California
; COUNTRY: United States of America
; ZIP: 94104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/011,366
; FILING DATE: 16-No. 6573003-2001
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/957,310
; FILING DATE: 23-OCT-1997
; APPLICATION NUMBER: US 08/329,154
; FILING DATE: 24-OCT-1994
; APPLICATION NUMBER: US 08/161,907
; FILING DATE: 02-DEC-1993
; APPLICATION NUMBER: US 07/985,321
; FILING DATE: 04-DEC-1992
; APPLICATION NUMBER: US 07/429,791
; FILING DATE: 31-OCT-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Ingolia, Diane E.
; REGISTRATION NUMBER: 40,027
; REFERENCE/DOCKET NUMBER: OPHD-01121
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 705-8410
; TELEFAX: (415) 397-8338
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2710 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 6:
US-10-011-366-6

Query Match 2.6%; Score 250.5; DB 4; Length 2710;
Best Local Similarity 18.1%; Pred. No. 8.9e-08;
Matches 387; Conservative 274; Mismatches 647; Indels 829; Gaps 101;
Qy 180 LSNFSAQSIKPLALTRSSDF--IAKLNQFNQDELWV-----YLEKPPDLALKA 227
Db 71 LSNLKDILKEVILTKNSNTSPVEK-----NLHFVWIGGEVSADIALEYTKQWADINA-EY 124
Qy 228 NLRLOTADSPFEGNLDVDPVYSF-----IRNPQ--NQEKWASDLN--QDQKT 271
Db 125 NIKLWDSAFVLNWLKKAIVESSTEALQLLEEELQNFQDNMKFYKKRMEFIYDROKR 184
Qy 272 VRLYRTEFS-----POAKTILKDY--KY-KDETFLSS-----IDLKASNG 309
Db 185 FINYKQINKPTVPTDIDIIKSHLVSEYNRDETVLSEYRTNSLRKINSNHGDIRAN-- 242
Qy 310 TSLFA-----NENDLKQOLDVLDVSDYFGQSEETITNSOVKRPVPASERS 356
Db 243 -SLFTEQELLNIYSOELLNRGNLAASDVRLLALKNFG----- 281
Qy 357 LKDRVKFKKDOQKPRIEKFSLY-EYDAL--SFYSQQLVSKPNSI----- 399
Db 282 -----VYLDVDMPLGIIHSDLFKTISRPSGIGLDRWEMIKLEAIMK 321
Qy 400 -KDLVNATLARNLRFSLGKYNFLFDDLAHL--DYFLVSKAKIKQSSITKFKLPIELPIKI 457
Db 322 YKKYINNYTSEN-----FDKLDQOLKDNFKLIIESKSEKSEIFSKL----- 362
Qy 458 SLKSSILGQOEENIKTLFEKEVTFKLDNPRDVEIEKAFGLLYPGVNEELEQARKAQRASP 517
Db 363 -----ENL-----NVSDLEIKIAFAL-----GSVINQALISQGSY 393
Qy 518 EKEKSKKGLKFSQOKEENSKAINNQE-----GLEEDDNITE-----RLPNSPI 562
Db 394 -----LTNLVIEQVKRYQFLQHLNPAIESDNNFTDTTKIFHDSLFSNATAENSMP 445
Qy 563 -----QYQENAGLGASPKPYMIK-----DVQNRYY-----LAKSOIQLIKAKDYTKLA 609
Db 446 LTKIAPYLQ-----VGFMPEARSTISLGGPGAVASAYYDFINLQENTIEKTLKASD----- 496
Qy 610 KLLSNRHTYNI SLRLKEQLFDVNPRIPSRDIKAKFVLDTKKNKYWQIYSSASPVFQN 669
Db 497 -LIEFKPPENNLSQLTEQ--EINSL--WSFQASAKYQFEKYVRDVTGSGLSBEDNGVDNF 551
Qy 670 KWSLFGYRYLLG-----LDPKQTIHELKVLGKAGLQFEGYENLPSPDFNLEDLK 719
Db 552 KNTALD-KNYLLNNKIPSNVVEAGSKNVYHVIQL-QGDDISYEATCNL---FSKNPKN 606
Qy 720 NTRIKTPLFSQKDNFKL-----SLDDFNYYDGEIKAPE-----FGLPLFL 760
Db 607 STIIQRNMNESAKSYFLSDDGESILELKY-----RIPERLKNKEKVKVTFIGHGDEFN 661
Qy 761 PKELARNSSNGSGSONS-----SPWQOEI-----ISQFKDQNLNQDQLAQFSTKI 807
Db 662 TSEFARLSVDSLNSNEISFLDTIKDISPKVNEVNLGCMFSDYDNVEETVPGKLLLSI 721
Qy 808 WEKIIIGDENEFDQNN-----RLQYKLLKDL---QESWINK-----TRDNYL 845
Db 722 MDKITSTLPDVNKNSTITIGANQVEYRINSEGRKELLAHSGKWINKEAAMS DLSKEYIF 781
Qy 846 WTYLGDKLVKPKNNLEAKFROISNLQELLTFATYSAALS-----NN----- 887
Db 782 FSDIDNKLKAKSKN-----IPGLASISEDIKTLLDASVSPDTKFLNNLKNLNISSIGD 836
Qy 888 WNYODSGAKSTIIFEEIAEL--DPKVKVKGVVYOLK-----PHYAIGFDD---NAG 936
Db 837 YIYYEKLEPVKNIIHNSDDDLIDEFNLENVSDLEYELKLNLDKEYLISFEDISKNN 896
Qy 937 KFNQEI-RSSRTIYLYKTSGK-----SKLEADTID 966
Db 897 TYSVRFINKSGESVYVETEKEIFSKYSBHITKEISTIKNSITDVGNNLNDIQLDHTS 956

Qy	967	QLNQAUVNAPLGLOSYFLDTERPGVQKLAATSLAVQ----	-----	1002
Db	957	QVN--TLNAFFIQSLIDYSSNKNVLDLSTSVKYAQLFSTGLNTIYDSIQLVNLIS	1014	
Qy	1003	-----	-----HKQKEKT-----	1009
Db	1015	NAVNDTINVLPITTEGIPIVSTILDGINLGAALKELLDEHDPLLKLEAKVGVLAINMS	1074	
Qy	1010	-----	-----LPKKLNDGYTLIHDKLKKPV--IPQISSS	1037
Db	1075	LSIAATVASIVGIGAEVTIPLPIAGISAGISPLVNN--LILHDKATSVVNYFNHLSSES	1132	
Qy	1038	-----	-----PEKDWFEGLKNQ-----	1063
Db	1133	KKYGPLKTEDDKILVPIDDLVISEIDFNNNNSIKLGTCTNILAMEGGSGHTVTGNI	1192	
Qy	1064	IIESPVSFTNFQEDADLDODGODDSRQ-----	-----GNNSLDNQEAGL	1102
Db	1193	PSISSHIPSLISYSAIGIETENLDFSKIMMLPNAPSRVFWMETGAVPGLRSLENDGTRL	1252	
Qy	1103	LKOKLAILLGNQFIQYQQNDKBEI-----	-----FEIINVEKVSELSFR-----	1146
Db	1253	LDSIRDLYPGKFWRFYAFPDYALITTLKPYVEDTNIKIKLDKDRNFIMPTTITNEIRNK	1312	
Qy	1147	LAKTLEDNKGKTIIRVL-----	-----SDEWMSL--IVNTTIKTPEMSAV	1183
Db	1313	LSYSFDGAGCTSYLLLSYSPISITNLSKODLWIFNIDNEVREISIENTIKKGLIKDQV	1372	
Qy	1184	PEVPDTKWVBOYDPRTPLAAKTFLVKFKDQIPVDGSGNISDK-----	-----WL	1228
Db	1373	LSKIDIN-----	-----ANKLII--GNQIYDFSGDIDNKRIVFLTCLDDKISLI	1416
Qy	1229	ASIPLVIHQWML-----	-----RLSPVVKTIIBRELGLK-----	1269
Db	1417	IEINLVAKSYSLLSGDKNYLISNLSNTIEKINTLGLDSKNIAVNYVTDSENNKYFGAISK	1476	
Qy	1270	OPOKKAVRKEEB-----	-----LETNTP-----	1295
Db	1477	TSQKSIITHYKDSKNILFEFYNDSTLEFNSKDFIAEDINVPKODINTITGKYVYDNN	1536	
Qy	1296	-----	-----PLTKAHLRTLNLVNDNPNYKIEDLVKNEAGHQDLAFSLRANNIKRLMNTPI	1349
Db	1537	SIDPSISLVKNQVKNGLYNESVYS--SYLDFVKNSDGH-----	-----NTSNFNNLFLDNI	1589
Qy	1350	TFADYNPFYYNEDWRSIDKYL-----	-----NNKG-----	1385
Db	1590	SF--WKLFGPENINF-VIDKYFTLVGTLNGLYVEFFICDNNKNIDIVFGWEKTS	1646	
Qy	1386	GGNGGS-----	-----GLIORLKN--NIKPTTFPALIALKDRN	1417
Db	1647	SGNGRNVVVEPIYNPDTCGEDI	1701	
Qy	1418	NTNL--SNYSKTIIMIKPK-----	-----YLVERSIGVPWSTGLDYGISEQTOKDGTSSSQ	1467
Db	1702	NTNYSNEYYPEIIVLNPNTTHKKVNNLDSSEPYKWSGEGDFILVRVLESNNKILQ	1761	
Qy	1468	Q---KGFDQDPTQALGLKNTYHGHKGL-----	-----SIRIFDPGNELEAK-----	1505
Db	1762	KIRIKGI-----	-----LSNTQSFNKMISDFKD IKKLSIGYIMSNFKSFENSELDRDHLG	1812
Qy	1506	IKDASNK-----	-----KGBEKLKSYDLFPKNVLYNEKKSPKIAKGTWNHHPDQKEYPNPNOKLP	1561
Db	1813	FKIIDNKTYYYDEDSKLVKGLININNSLFYDPDPIEFNLVTGWQTI--NGKKY-----	-----	1862
Qy	1562	ENYLNVLNQWKVTLY--NSSDFITN-----	-----LFVEPEGSDRGSGTKLKQVIOKQV	1610
Db	1863	--YFD--INTGAALTSYKIIINGKHIFYNNDGMQGVGFKGPDGPEYPAPANTQ-----	-----	1911
Qy	1611	NNYNAD-----	-----WGSALVT-----	1653
Db	1912	NNNTEGQAIYQSFRLTLNGKKYFYFDNNNSKAVTGWRIINNEKYYFNPNNA--IAAVGLQVI	1970	
Qy	1654	EDNTKLIAPN---ITQWPNPISGSKEFYKPTVPFGN	1687	

[illegible]

Db 185 FINYKQINQKPTVPTIDDIKSHLVSEYNRDETIVLESYRTNSLRKINSNHGIDIRAN-- 242
QY 310 TSLFA-----NENDLKQDLDVLLDVSYFGQSETITTSNQVKVPASERS 356
Db 243 -SLFTQELLNIYSQELLNRGNLAASDIVRLALKNFGG----- 281
QY 357 LKDRVFKDQKQPRIEKFSLY--BYDAL--SFYSQQLVSKPNSI----- 399
Db 282 -----VYLDVMDLPGIHSDLFKTISRPSISGLDRWEMIKLEAIMK 321
QY 400 -KDLVNATLARNLRFSLGKNFPLFDDLAHL--DYFVLVSKAKIKQSSITKLFIELPIKI 457
Db 322 YKYYINNYTSEN-----FDKLDQQLKDNFKLITTESKSEIIFSKL----- 362
QY 458 SLKSSILGDQEPNIKTLFEKEVTFKLDNFRDVEIERKAFGLLYPGVNEEELQAKAQRASF 517
Db 363 -----ENL-----NVSDLEIKIAFAL-----GSVINQALISQGSY 393
QY 518 EKESKKGKLFQSOQKEENSKAINQOE---GLEEDDNITE-----RLPENSP 562
Db 394 -----LTNLVIEQVKNRYQFLNQHLPALIESDNNFTDTKIFHDSLFSNATAENSMF 445
QY 563 -----QYQENAGLQASDPKPYMIK---DVQNRYY-----LAKSOIELIKAKDYTKLA 609
Db 446 LTKIAPYLO-----VGFMEPEARSTISLGGPAGAYAYDFINLOENTIEKTKASD----- 496
QY 610 KLASNRHTYNIURLKEQLFDVNPRIPESSRDIEKAKFVLDTKTEKNYQWQIYSSASVPQN 669
Db 497 -LIEFKPEPNLSQLTEQ---EINSL--WSPDQASAKYQEKYVRYDTGGSLSDNGVDNF 551
QY 670 KWSLFGYRYLLG-----LDPKQTHLHVKLQKAGLQFEGENLPSDFNLEDLK 719
Db 552 KNTALD--KNYLLNKNKIPSNVVEAGSKYVHYIIQL--QGDDISYEATCNL---FSKNPKN 606
QY 720 NIKIKTFLFQKQNFKL-----SLDDPNYVDEIKAPE-----FGLPLFL 760
Db 607 SIIIQRMNBSAKSYFLSDGSEILENKY-----RIPERLKNKVKVTFIIGHGKDFN 661
QY 761 PKELRRNSSGSGQSN-----SPWEQEI-----ISQFKDQNLNQDLAQFSTKI 807
Db 662 TSFARLSDVLSNEISFLDTIKDISPKVNEVNLGCNMFSDNVEETYPCKLLLSI 721
QY 808 WEKIIIGDENFQDN-----RLQYKLLKDL---QESWINK-----TRDNLY 845
Db 722 MDKITSTLPDVNKNSTIGANQVEVRINSGRKELLASHGKWINKEEAIMSDLSKEYIF 781
QY 846 WTVLGDKLKVPKNLEAKERQISNQLLELTAFYTSAAIS-----NN----- 887
Db 782 FDSIDNKLKAKSKN-----IPGLASISEDIKTLILDASVSPDTKFTLNNLKLNISSIGD 836
QY 888 WNYQDSGAKSTIIFEBIAEL--DPKVEKVGADVYOLK-----PHYAIGPDD---NAG 936
Db 837 YIYEKLEPVKNIIHNSIDDLIDEFNLENVDELVELKLNLDKYLISFEDISKNS 896
QY 937 KPNQVI--RSSRRTIYLKTSK-----SKLEADTID 966
Db 897 TYSVRPFINKSGESVYVETEKEIFSKYSEHITKEISTIKNSIITDVGNLLDNIQLDHTS 956
QY 967 QLNQAVKNAPLGLQSFYLDTERGVQKLATSVAQ----- 1002
Db 957 QVN--TLNAAFFIQSLIDYSSNKDVLNLDLSTSVKQVLAQLFSTGLNTIYDSIQLVNLIS 1014
QY 1003 -----HKQKEKT----- 1009
Db 1015 NAVNDTINVLPITTEGIPVISTILDGINLGAAILKEILLDEHDPLLKLEAKVGVLAINWS 1074
QY 1010 -----LPKLLANDGYTLIHDKLKXKPV--IPOISS 1037
Db 1075 LSAATVASIVGIGAEVTIFLLPIAGISAGIPSLVNN--LILHDKRATSVVNYFNHLS 1132
QY 1038 -----PEKOWFEKLNQON-----QSQNVNVSTFGS- 1063

Db 1133 KKGPLKTEDDKILVIPIDDLVISEIDFNWNSIKLGTNCILAMEGGSGCHTGTGNIDHFFSS 1192
QY 1064 -IIESPYSTNFQEDADLDQDQDSDRQ-----GNSLNDQEAGL 1102
Db 1193 PSISSHIPSLSIYSIGAIBETENLDFSCKIMLNPASRVFWMETGAVPGLRSLENDGTRL 1252
QY 1103 LKQKLAAILGNQFIQYQQNDKEIE-----FPIINVEKVSLSFR-----VEPK 1146
Db 1253 LDSIRDLYPGKPYRWFYAFDYAITTLKPVYEDTNIKIKLDKDRNFIMPTITTNIRNK 1312
QY 1147 LAKTLEDNGKTRVL-----SDETMSL--IVNTTIEKTPMSAV 1183
Db 1313 LYSYFDGAGTYSLLSSYPISNTINLSKDDLIWIFNIDNEVREISIENTIKKGLIKDV 1372
QY 1184 PEVFDTKWVEQDPRTPLAAKTFVKFKDQIPVDSGNISDK-----WL 1228
Db 1373 LSKIDIN-----KNKLII---GNQTIIDFGSDIDNKDRYIFUTCELDDKISLI 1416
QY 1229 ASIPLVIHOOML-----RLSPVVKTRRELGLK-----TEOQQQQQQQQQQ 1269
Db 1417 IBINLVAKSYSLLSGDKNYLISNLTIEKINTLGLDSKNIAYNVYDESNNKYFGAISK 1476
QY 1270 QPKKAVRKEEB---LETYNP-----KDEPNILN----- 1295
Db 1477 TSQKSIHYKDSKNILFEYNDSTLEFNSKOFIAEDINVFMKDDINTITTKYVVDNNTDK 1536
QY 1296 -----PTKAHRLTSLNVLNNDPNYKIBDLKVIKNEAGDHQALFASLRANNIKRLMTP 1349
Db 1537 SIDFSISLVSKNQKVNGLYLNEYSYS--SYLDFVKNSDGH-----NFSNFWNLFLDNI 1589
QY 1350 TPADYNPPYPYNEDWRSIDKYL-----NNKG-----NVSSHQQOAA 1385
Db 1590 SF--WKLFGFENINF--VIDKYFTLVGKTNLGVFEFICDNNKNIDYFGEWKTSSSKSTIF 1646
QY 1386 GGNQGS-----GLIQLRNLK--NIKPETFTPALIALKDRN 1417
Db 1647 SGNRNVVVEPIYNPDGTGEDISTSLDFSVEPLYGIDRYINKVLIAPDLYT--SLINI----- 1701
QY 1418 NTNL--SNYSDKIIMIKPK-----YLVBSRISGVPMWSTGLDGYIGSQTGDTSSSSQ 1467
Db 1702 NTNYSNEYYPEIIVLPNPTFHKNVINLDSSEFYKWKSTEGSDFLVRYLSESNKKILQ 1761
QY 1468 Q---KGFDDQFQALGLKNTYHGKGL-----SIRIFDPGHLEAK----- 1505
Db 1762 KIRIKGI-----LSNTQSFNKSIDFKDKIKLSLGYIMNSFKSFENSELDRHLG 1812
QY 1506 IKDASNK---KGEKLLKSYDLFKNYLNEYKKGPKIAKGWNTNIHPDOKEYPNPQKLP 1561
Db 1813 FKIDNKTYVDEDSKLVKGLININNSLFYDPDIBENLVTHQWTI---NGKYY----- 1862
QY 1562 ENYLNVLNQPWKVTLY---NSSDFITN-----LFVEPEGSDRGSGTKLKQVIOKOV 1610
Db 1863 --YFD--INTGAALTSYKIINGKHFFYFNNDQVNLGVKPGDGFYFAPANTQ----- 1911
QY 1611 NNNYAD---WGSAYLT-----FWVDKN-----IINQPNVITANADVFIDKVKEL 1653
Db 1912 NNNIEQAIYVQSKELTNKKKYFYDNNNSKAVTGWRIINNEKYFNPNNNA--TAAVLQVMI 1970
QY 1654 EDNTKLIAPN---ITQWPNISGSKKPYKPTVFFGN 1687
Db 1971 DNNKYFNPDTAILSKGWOTVNGSYFYFDTDTAIAFN 2007

RESULT 25

US-08-603-753D-4

; Sequence 4, Application US/08603753D

; Patent No. 5891857

; GENERAL INFORMATION:

; APPLICANT: HOLT, JEFFREY T.

; APPLICANT: JENSEN, ROY A.

; APPLICANT: PAGE, DAVID L.

; APPLICANT: KING, MARY-CLAIRE

; APPLICANT: SZABO, CSILLA I.

APPLICANT: JETTON, THOMAS L.
 APPLICANT: ROBINSON-BENION, CHERYL L.
 APPLICANT: THOMPSON, MARILYN E.
 TITLE OF INVENTION: CHARACTERIZED BRCA1 AND BRCA2
 TITLE OF INVENTION: PROTEINS AND SCREENING AND THERAPEUTIC METHODS BASED ON
 TITLE OF INVENTION: CHARACTERIZED BRCA1 AND BRCA2 PROTEINS.
 NUMBER OF SEQUENCES: 29
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: ARLES A. TAYLOR, JR.
 STREET: SUITE 1401, UNIVERSITY TOWER, 3100 TOWER
 STREET: BOULEVARD
 CITY: DURHAM
 STATE: NORTH CAROLINA
 COUNTRY: USA
 ZIP: 27707

COMPUTER READABLE FORM:
 MEDIUM TYPE: Diskette, 3.50 inch, 800 kb storage
 COMPUTER: IBM PC/XT/AT compatible
 OPERATING SYSTEM: Windows 3.1
 SOFTWARE: WORD PERFECT 6.1 and ASCII
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/603,753D
 FILING DATE: 20 FEB 1996
 CLASSIFICATION: 514
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: U.S. 08/373,799
 FILING DATE: 17 JAN 1995
 ATTORNEY/AGENT INFORMATION:
 NAME: ARLES A. TAYLOR, JR.
 REGISTRATION NUMBER: 39,395
 REFERENCE/DOCKET NUMBER: 1242/2
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (919) 493-8000
 TELEFAX: (919) 419-0383
 TELEX:

INFORMATION FOR SEQ ID NO: 4:

SEQUENCE CHARACTERISTICS:

LENGTH: 3418

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: unknown

MOLECULE TYPE: protein

HYPOTHETICAL: no

ANTI-SENSE: no

ORIGINAL SOURCE:

ORGANISM: Homo sapiens sapiens

INDIVIDUAL ISOLATE:

DEVELOPMENTAL STAGE: adult

TISSUE TYPE: female breast

CELL TYPE: normal breast tissue

CELL LINE: HMEC

ORGANELLE: no

FEATURE:

NAME/KEY: BRCA2 protein

LOCATION: 1 to 3418; Genbank locus HSU43746

IDENTIFICATION METHOD:

OTHER INFORMATION: BRCA2 protein has a negative

OTHER INFORMATION: regulatory effect on growth of human mammary cells.

PUBLICATION INFORMATION:

AUTHORS: Wooster, R. et al.

TITLE: Identification of the breast cancer

TITLE: susceptibility gene BRCA2

JOURNAL: Nature

VOLUME: 379

PAGES: 789-792

DATE: 1995

RELEVANT RESIDUES IN SEQ ID NO: 4: granin box

RELEVANT RESIDUES IN SEQ ID NO: domain at amino acids 3334-3344

US-08-603-753D-4

Query Match

Best Local Similarity 2.6%; Score 250.5; DB 2; Length 3418;

Matches 387; Conservative 280; Mismatches 716; Indels 693; Gaps 96;

85 EFTDFVSKFDFLTNNGRVTLEIPKQVWVIEBSPB-----DKERF 126
 1126 EFTQF-RKPSYILQ--KSTFEVPEWQWTLKTTSECRDADLHVIMNAPSIGQVDSKQF 1182
 127 RLGFHLKEKLEDGNIQAQSAKFYLLPLDMPKALGOYSYIVDN-----171
 1183 EGTVEIKR-----KFAGLLKNDCKNSAG--YLTDENEVGRFPYSAHGTLK 1227
 172 -----FNNLIHPLSNFSAQSIKPLALTRSS--DEIAKLQNFQNODELWVYL 216
 1228 NVSTEALQKAVKLFSD--IENISEETSAEVHPISLSSCKCHDSVWSFKIENHNDKT VSE 1285
 217 EKFFDLEALKANIRLOTADSPFEKGNLVDPVVYSFIRNPQO-----258
 1286 KNNKQILQNNIEMT-----GTFVEITEYKRNTEENEDNKYTAASNSHNLEFDG 1338
 259 -----KEMASDLNQDQKTVRLYLRTFPSPQAKTILKDYKYKDTFTFSSIDLK- 305
 1339 SDSSKNDTVCIHKDETDLLFTDQHNICLSLGGQFMKEGNTQIKR-DLSDLTFLLEVAKAQE 1397
 306 -----ASNGTSLFA--NENDLKOQLOVDLDVSDYFGQSGSETITSNSQKVPVPSERSLK 358
 1398 ACHGNTSNKEQLTATKTEQNIKO-----FETSD-----TFQTAGSKNISVAKSEFN 1444
 359 DRVKFKKQDQKPRIKESL-----YEYDALSPY-----SQL 389
 1445 KIVNF-FDQRPBELHNSLSELHSDIRKKNMILSYEETDIVKHKLKESVPVGTGNQL 1503
 390 QELVSKPENSIKDLVNATL-----ARNLRFSLGKYNPLF-----423
 1504 VTFOGQPERDEKIEPTLLGFHTASGKKVIAKESLDKVNKLDFEKEQGTSEITFSFHOW 1563
 424 -----DLASHLDYFVLSKAKIKQSSITKK--LFIELPIKLSLSSIIIGDQ 467
 1564 AKTKYREACKDLELACETIETAAPKCKEQMNSLNDKNLVSETVPPKLLSDNLCRQ 1623
 468 EPNIKTLFEKEVTPKLDNFRDVEIEKAFGLLYPCWNEE-----LEQAKAORASPEKES- 522
 1624 TENLKT--SKSIFLKVKVHENVKEKETAKSPATCYTNQSPYSVIENSALAFYTCRSKRTSV 1681
 523 -----KRGKLE--FSQKKEENSKAINNOGLEEDDNITERLPENSPYOQOENAGLG 572
 1682 SQTSLLEAKKWLREGIFDGGQPERINTA-----DYGVNLYENNSNSTIAEN-----1727
 573 ASPDKPMIKDVQORYLAKSQOELIKADYTKLAKLLSNRHTYNI SLRLKQLFDVN 632
 1728 ---DKNHL---SEKQDTYLSNS-----MSNSYSYH-----SDEVYNS 1760
 633 PRIPSSRDIEKAKFVLDKTEKNKYQIYSSASPVFQNKWLSFGYRYVLLGLDPKQTIHEL 692
 1761 GYLSKNKLDLSGIEPLVKNVEDOK-----NTSFKVINSVKDANAY-----PQTVNE- 1806
 693 VKLQKQKAGLOPEGYENLPDFNLBDLKNIRIKTLPFSQKONFKLSLLDFNNYVGEIKAP 752
 1807 -----DICVEEL--VTSSSPCKNKAIAKLSISNSNF---EVGPP 1842
 753 EFGPLPLFLPKELRRNSNSGSSQNSNSPWEDEIISQPKD-----QNLNQDQLA 801
 1843 AFRI-----ASGKIRLCS---HETIKVKVDIFDTSFKVIKENNENKSKIC 1885
 802 QFSTKIM-----EKIIGDENEFQDNRLQVKKLQKESWINKTRDNLYWYTLGD 851
 1886 Q--TKIMAGCYEALDDSDILHNSLNDDECSMHSHKVPADIQSEILQHNQM--SGLEK 1941
 852 KLKVKPKN-NLEAKPRQISNLQELLTAPYTSAAALSNMNNYVQDSGAKST-----899
 1942 VSKISPCDVSLSDTICKCSIGKL---HKSVSANTCGIFSTASGKSVQVSDASLQVAR 1997
 900 IIFEIAELDPKVKVKV-----GADVQLKFHVAIGFDD---NAGFNQSVISSSRRTI 950
 1998 QVFSIEDSTKQVFSKVLFSKNSHSDQLTRENTAIRTPHEHLISQKGFSVNVNSSFASFSG 2057

Qy	951	YLTSGSKSLEADTIDQNOAVKNAPJGLQSF-----YLDTERFGVFOKLATSLA	1000
Db	2058	FSTASGK--QVSILSSHLHKVGV--LBEEFDLIRTEHSLHSPTSQRQV-----	2102
Qy	1001	VQHKQKETTLPKLLNNGYTLIHDKLAKKPVIPQISSSPKDWEGKLNONGSQNVNVT	1060
Db	2103	-----SKIIPRVDKRPEHCNVSEMEKTSCKPKLS-----NNLNVEGGSSENN---	2146
Qy	1061	FGSIISEPYFTNFQEDADLDQDQDDSRQGNNSLWQEAAGLLKQKLAILLGNQFIQYYQ	1120
Db	2147	HSIKVSPYLS-QFQD-----KQQLVLGTQKVSLENHIVLHGKEQ	2184
Qy	1121	QNDKEIEFEINVEKVSSELSFRVEFKLAKT-----LENGKTI-IR	1159
Db	2185	ASPKNVKWEIGKTEFTFDGVPVKTNIEVCSTYSKDSENYFTEAVEIAKAFMEDELTDSK	2244
Qy	1160	VLSDETMSL-----IWNVTIEK-----TPEMSAVPEVDFDKWBOYDPRTPLAATK	1206
Db	2245	LPSHATHSLTCCPENEEMWLSNSRIGKRGEPILVLGEPISKNNLLNEFDRIENQEK-	2303
Qy	1207	FVLKFDQDIPVDSGNSIDKWLASIPILVHQQMLRLSPVV-----KTIRELGLKTEQOQQQ	1262
Db	2304	--LKASKSTP--DGTIKDR-----RUFMH--VSLEPIICVPFRTTKE-----	2340
Qy	1263	QOQOQOQOQKAVKKEELETYNPKDEFNLIPLTKAH-----RLTL-----SNL-VNNDPN	1314
Db	2341	--ROEIQNPNTA-----PGQEF-----LSKSHLYEHLTLEKSSSNLAVSGHPF	2382
Qy	1315	YKTEDLKVIKNEAGDHQALSLRANNIKRLMNTPIITPADYNPFYFNEDWRSIDKYLNN-	1373
Db	2383	QOVS-----ATRNEKWRHLITGRPTKVFVPFKTSHPHRVEQCVRNI	2426
Qy	1374	-----KGNVSHQOQAAGNGSGLIQRLNKNIKE-----TTPA-----LIAKDR	1416
Db	2427	NLENRQONIDGHGSDSKNKINDNEIHQFNKNSNQAAAVFTKCEEBPLDLI-----	2481
Qy	1417	NNTNLSNYSD-KIIMIKPK-----VLVERSIGVPWSTGLDGYIGSEOTKDGTS	1463
Db	2482	--TSLQWARDIQMRIKKQRQVFPQPGSLYIAKTST-LP-RISLKAAVGGVPSACSH	2537
Qy	1464	SSSQQGFQDQFIQALGLKNTE-----YHGK-----LGLSIRIFDPGNELAKIKDASN	1511
Db	2538	KQLYTVGSXHCIK-INSKNAESFQPHTEYFGKESLWTCKGIQLADGG--WLIFSDNG	2593
Qy	1512	KGBEELKLSYDIFKNYLYNEYKKS PK-IAKGWNIHPDOKEFPNPNQKI-----PEN	1563
Db	2594	KAGKEE-----FYRALCPTPGVDPKLISRIWVYNH-----YRWIITWLAAMECAFPE	2641
Qy	1564	YLNVLNLHOPWKVTLNNSDDITNLFVEPEGSGRSGTKLQVKQKOVNNNYADWGSAYLT	1623
Db	2642	FANRLS-PRVLLQLYRYDTEI-----DSRRSAIKNIWER-----DDTAAKTL	2686
Qy	1624	FWYDKNIITQPNVITANIDVFIKDYKELEDNTKLIAPNITQWPWNISGSKKFKYKPTV	1683
Db	2687	VLCVSDIIS-----LSANISSETSNKTS-ADTKQVAIIBELTDGW-----	2725
Qy	1684	FFGNWENESSMNSOACTPTWEKIREGFAQLAKSSFDOK-----	1723
Db	2726	-----YAVKAQLDPPPLLALVKNG-----RLTVGQKIILHGAEVLGSPDACPTEA	2770
Qy	1724	--TRTFVLTNAPLPLWKYGLFGQNGPNEFKTDWRLVFDONDN-----QIAALRVQEQDR	1777
Db	2771	PESIMULKISANSTPARWYTKLGFPPDPFPPLPLSLFLSDGGNGVCVDVIOQRYPIQR	2830
Qy	1778	PEKSS-----EDDKQKQWIKFKVVIPEEMF	1802
Db	2831	MEKTSGLGYIFRNEREBEKEAAKYVEAQQRKLEALF	2866

GENERAL INFORMATION:
 APPLICANT: HOLT, JEFFREY T.
 APPLICANT: JENSEN, ROY A.
 APPLICANT: PAGE, DAVID L.
 APPLICANT: KING, MARY-CLAIRE
 APPLICANT: SZABO, CSILLA I.
 APPLICANT: JETTON, THOMAS L.
 APPLICANT: ROBINSON-BENION, CHERYL L.
 APPLICANT: THOMPSON, MARILYN E.
 TITLE OF INVENTION: CHARACTERIZED BRCA1 AND BRCA2
 TITLE OF INVENTION: PROTEINS AND SCREENING AND THERAPEUTIC METHODS BASED ON
 TITLE OF INVENTION: CHARACTERIZED BRCA1 AND BRCA2 PROTEINS.
 NUMBER OF SEQUENCES: 29
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: ARLES A. TAYLOR, JR.
 STREET: SUITE 1401, UNIVERSITY TOWER, 3100 TOWER
 STREET: BOULEVARD
 CITY: DURHAM
 STATE: NORTH CAROLINA
 COUNTRY: USA
 ZIP: 27707
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Diskette, 3.50 inch, 800 kB storage
 COMPUTER: IBM PC/Xt/AT compatible
 OPERATING SYSTEM: Windows 3.1
 SOFTWARE: WORD PERFECT 6.1 and ASCII
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/09/099,753
 FILING DATE:
 CLASSIFICATION:
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: 08/603,753
 FILING DATE: 20 FEB 1996
 APPLICATION NUMBER: U.S. 08/373,799
 FILING DATE: 17 JAN 1995
 ATTORNEY/AGENT INFORMATION:
 NAME: ARLES A. TAYLOR, JR.
 REGISTRATION NUMBER: 39,395
 REFERENCE/DOCKET NUMBER: 1242/2
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (919) 493-8000
 TELEFAX: (919) 419-0383
 TELEX:
 INFORMATION FOR SEQ ID NO: 4:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 3418
 TYPE: amino acid
 STRANDEDNESS: single
 TOPOLOGY: unknown
 MOLECULE TYPE: protein
 HYPOTHETICAL: no
 ANTI-SENSE: no
 ORIGINAL SOURCE:
 ORGANISM: Homo sapiens sapiens
 INDIVIDUAL ISOLATE:
 DEVELOPMENTAL STAGE: adult
 TISSUE TYPE: female breast
 CELL TYPE: normal breast tissue
 CELL LINE: HMEC
 ORGANELLE: no
 FEATURE:
 NAME/KEY: BRCA2 protein
 LOCATION: 1 to 3418; Genbank locus HSU43746
 IDENTIFICATION METHOD:
 OTHER INFORMATION: BRCA2 protein has a negative
 OTHER INFORMATION: regulatory effect on growth of human mammary cells.
 PUBLICATION INFORMATION:
 AUTHORS: Wooster, R. et al.
 TITLE: Identification of the breast cancer
 TITLE: susceptibility gene BRCA2
 JOURNAL: Nature
 VOLUME: 379
 PAGES: 789-792

Mon Nov 21 16:44:20 2005

DATE: 1995	
RELEVANT RESIDUES IN SEQ ID NO: 4: granin box	
RELEVANT RESIDUES IN SEQ ID NO: domain at amino acids 3334-3344	
US-09-099-753-4	
Query Match	2.6%; Score 250.5; DB 3; Length 3418;
Best Local Similarity	18.6%; Pred. No. 1.3e-07;
Matches 387; Conservative 280; Mismatches 716; Indels 693; Gaps 96;	
85 EFTDFVSKFDFLTNGRTVLEIPKQYQVISEFSE	126
1126 EFTQF-RKPSYILO--KSTFEVPEVQMTILKTSEECDADLHVIMNAPSIGQVDSKQF	1182
127 RLGFHLKEKLEDGNIASQATFYIYLLDMPKALQYSIYVDN	171
1183 EGTVEIKR-----KFGALLKNDCKNSAG--YLTDENEVGRFGYSAHTKL	1227
172 -----FNNLIHPLSNFSAQSIKPLALTRSS--DFAKLNQFNQDELWYL	216
1228 NVSTEALQAVKLFS--IENISETSAEVHIPISLSSCKCHDSVSMFKIENHNDKTVSE	1285
217 EKFFDLKALANIRLOTADFSEKGNLVDPFVYFIRNPQO-----	258
1286 KNNKCOLLQNNIEMT-----GTFVEITEYKRNTEENEDNKYTAASRSHNLBFDG	1338
259 -----KEWASDLNOQKTVRLYLRTFSPQAKTILKDYKDETFELSSIDLK- 305	
1339 SDSSKNDTVCIHKDETDLFTDQHNICLKSGQFMKEGNTQIKE-DLSDLTFLEVAQAQE 1397	
306 -----ASNGTSLFA--NENDLKQDLVDLLVDVFGGQSITITSNSQVQVPVAPASERSLK 358	
1398 ACHGNTSNKEQTATKTSQNIKD-----FETSD-----TFTQASGKNIISVAKELFN 1444	
359 DRVFKKQKQPRIKFSL-----YEYDALSFY-----SQL 389	
1445 KIVNF-FQKPELHNFSLNSHSDIRKNKMDILSYEETDIVKHILKESVPVGTGNQL 1503	
390 QELVSKPNSIKDLVNATL-----ARNRPSLGRKNFLFD-----423	
1504 VTFQOQPERDEKIEPTLLGHTASGKVKYAKESLDKXNLFDEKEQGTSEITFSFHOW 1563	
424 -----DLASHLDYVFLVSKAKIKOSSITKK--LFIELPKISLKSSILGDQ 467	
1564 AKTKYREACHOLEACETIETIATAPKCKEMONSUNNDKNULVSIETVPPKLLSNLCRQ 1623	
468 EPNIKTLPEKVTFLKDNPRDVEIEKAFGLLYPGVNEE-----LEQAKAQRASFEKES- 522	
1624 TENLKT--SKSIFLKVKVHVENKETAQSPATCYTNQSPYSVIENSALAFYTSCSRKTSV 1681	
523 -----KXGLKE--FSQCKEENSKAINNQGLEDDNITRLPENSPIQYOQENAGLG 572	
1682 SQTSLLEAKKWLREGIFDQPERINTA-----DYVGNLYENNNSNSTIAEN-----1727	
573 ASPDKPYMIKDQVQRYYLAKSQIQELIKAKDYTKLAKLLSNRHTYTNISLRLEKOLFVN 632	
1728 ---DKNHL---SEKQDTYLSNS-----MSNSYTH-----SDEVYND 1760	
633 PRIPSRDIEKAFVLDTKKNKYMOIYSSASPVFONKWSLFGYRYLLGLDPKQTIHEL 692	
1761 GYLKRNKLSGIEPVLKVEDQK---NTSFSKVINVKDANAY-----PQTVNE- 1806	
693 VKLGQKAGLQFEGYENLSPDFNLEDLKNIRITKPLFSQKDNFKLLDPNNYDGEIKAP 752	
1807 -----DICVEEL--VTSSSPCKNKAIAKLSISNSNF---EVGPP 1842	
753 EFGPLPLPKELRRNSGGGQNSNSPWEQELISQFKD-----QNLSDQOLA 801	
1843 AFRI-----ASGKIRLCS---HETIKVKDKIFTDSFSKVTKENNENKSKIC 1885	
802 QFSTKIW-----EKIIGDENEFQDNNRLQYKLLKDLQESWINKTRDNLWYTLGD 851	
1886 Q---TKIMAGCYEALDDSEDLILNSLNDDECSMHSKVFADIQSEELIQHNQNM--SGLEK 1941	

QY	852	KLKVKPKN-NLEAKFRQISNLQELLTAFYTSAAALNNMNYQDSGAKST-----	899
DB	1942	VSKISPCVSLFSDICKCSIGKL-----HKSVSANTCGIFSTASGVSQVSDASLQANAR 1997	
QY	900	IIPETIAELDPKVKKEV-----GADVYQLKFHYALGPD-----NAGKNEQEVIRSSRTI 950	
DB	1998	QVPSIEDSTKQVSKVLFKSNEHSDQLTRBENTARTPEHLISQKGFSTNVNVSASFSG 2057	
QY	951	YLKTSCKSKLEADTIDOLNAQVKNAPLGLQSF-----YLDTERFGVFOKLATSLA 1000	
DB	2058	FSTASGK-----QVSLSESLHKVGV---LEEDPLIRTEHSLHVSPTSQRNV-----2102	
QY	1001	VQHKQEKTLPKKLNNDGYTLIHDKLKXPIVPOISSPEKDMFEGKLNQNGQQQNVNST 1060	
DB	2103	-----SKILPRVDKRNPEHCNVSEMEKTSCKEFKLS-----NNLNVGGSENN---2146	
QY	1061	FGSIIESPSTNYFOEDADLDQDQDDSRQNNSLONQEAAGLAKQKLAILLGNQFIQYQ 1120	
DB	2147	-HSIKVSPYLS-QFQD-----KOOLVLGTVKSVLNVENIHVLGKEQ 2184	
QY	1121	QNDKEIEFEIINVKVSELSFRVEFKLAKT-----LEDNGKT-IR 1159	
DB	2185	ASPRNVMELGKTETTSFDPVKTNIEVCSTYSKDSYNYFETEAVEIAKAFMEDDELTDK 2244	
QY	1160	VLSDETMSL-----IVNTTIEK---TPMSAVPEVDFTKWVQYDPRTPLAAKTK 1206	
DB	2245	LPSHATHSLFTCPENEMVLSNRSRIGRGEPLILVGEPSIKRMLLNEDPRIIENQES- 2303	
QY	1207	FVLKFKDQIPVDSGNSIDKWLASIPLVIIHQMLRLSPVV-----KITRELGLKTEQOOQQ 1262	
DB	2304	--LKASKSTP--DGTIKDR--RLFMMH--VSLEPITCVPFRTTKE-----2340	
QY	1263	QOQOQOQOQKAVRKEELETYPNDEFNILNPLTKAH--RLTL-----SNL-VNNDPN 1314	
DB	2341	--ROEIQNPFTA-----PGQEP-----LSKSHLYEHLTEKSSNLSAVSGHPF 2382	
QY	1315	YKIEDLKVIKNEAGDHLAFSLRANNIKRLMNTPIITPADYNNPFYNNEDWRSIDKYLN- 1373	
DB	2383	YQVS-----ATRNEKMRHLITGRPTKVFVPPPKTSHFHRVEQCVRN 2426	
QY	1374	-----KGNVSHQQAAGNQGSLIQRINKIKPE-----TTPA-----LIALKDR 1416	
DB	2427	NLEENRQKQNDHGSDSKNKNINDNEIHQFNKNNSNQAATAVFTKCEEPFLDLI-----2481	
QY	1417	NTNLSNYSO-KIIMIKPK-----YLVERSIGVPMSTGLDYGISEQTKDQTS 1463	
DB	2482	--TSLQWARDIQDMRIKKKORQVFPQPGSILYLAKTST-LP-RISLKAAGVQVPSACSH 2537	
QY	1464	SSSQKGFQDQFIQALGLKNT-----YHGK-----LGLSIRIFDPGNELAKIKDASN 1511	
DB	2538	KQLYTYGVSKHCIK-INSKNAESFQHTEDYFGKESLWTGKGIQLADG---WLIPSNDG 2593	
QY	1512	KKGBEKLKSYDLFQVLYNEVEKSPK-IAKGWNIHDPQKEYPNPQKL-----PEN 1563	
DB	2594	KAGKEE-----FYRALCDTPGVDPKLISRIWYNH-----YRWITKLAAAMECAPPKE 2641	
QY	1564	YLNVLNQPKVLYNSDDFITNLFVEPEGDRSGTKLKOVIQOVNNVADWGSAYLT 1623	
DB	2642	FANRCLS--PERVLLQKYRYDTEI-----DRSRSAIKKIMER-----DDTAATKL 2686	
QY	1624	FWYDKNIITNQPNVTIANIADVFIKDVKELEDNTKLIAPNITQWPNPISGSKFYPKPTV 1683	
DB	2687	VLCVSDIIS-----LSANISSETSNKTSS--ADTKQVAILIETDGM-----2725	
QY	1684	FFGNWENSSNSMGQAQPTTWEKIREGALQALXSSFPQK-----1723	
DB	2726	-----YAVKAQLDPPLLAVLKNQ-----RLTVGQKILHGAELVSGSPDCTPLEA 2770	
QY	1724	--TRTFVLTWAPLPLWKYGPLGQNGFNFKTQDWRLVFQNDN-----QIAALRVQEOQR 1777	
DB	2771	PESLMKISANSTRPARWYTKLGFPPDPRFPPLPLUSSLFGSGNVCVGVIIQRIAYPIQR 2830	
QY	1778	PEKSS-----EDKDKQKWKFKVWIPEEMF 1802	

Db 2831 MEKTSGLYIFRNEREBEKAAYVEAQQRLEALF 2866
||:| |::: |::| |:
RESULT 27
US-08-986-106-4
; Sequence 4, Application US/08986106
; Patent No. 6177410
; GENERAL INFORMATION:
; APPLICANT: HOLT, JEFFREY T.
; APPLICANT: JENSEN, ROY A.
; APPLICANT: KING, MARY-CLAIRE
; APPLICANT: STEINER, MITCHELL S.
; APPLICANT: ROBINSON-BENION, CHERYL L.
; APPLICANT: THOMPSON, MARILYN E.
; TITLE OF INVENTION: THERAPEUTIC METHODS FOR
; TITLE OF INVENTION: PROSTATE CANCER
; NUMBER OF SEQUENCES: 26
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: ARLES A. TAYLOR, JR.
; STREET: SUITE 1401, UNIVERSITY TOWER, 3100 TOWER
; STREET: BOULEVARD
; CITY: DURHAM
; STATE: NORTH CAROLINA
; COUNTRY: USA
; ZIP: 27707
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.44MB storage
; COMPUTER: IBM PC/XT/AT compatible
; OPERATING SYSTEM: Windows 3.1
; SOFTWARE: WORD PERFECT 6.1 and ASCII
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/986,106
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/603,753
; FILING DATE: 20 FEB 1996
; ATTORNEY/AGENT INFORMATION:
; NAME: ARLES A. TAYLOR, JR.
; REGISTRATION NUMBER: 39,395
; REFERENCE/DOCKET NUMBER: 1242/3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (919) 493-8000
; TELEFAX: (919) 419-0383
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3418
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: unknown
; FEATURE:
; NAME/KEY: BRCA2 protein
; PUBLICATION INFORMATION:
; AUTHORS: Wooster, R. et al.
; TITLE: Identification of the breast cancer
; JOURNAL: Nature
; VOLUME: 379
; PAGES: 789-792
; DATE: 1995
; RELEVANT RESIDUES IN SEQ ID NO: 4: granin box domain at
; RELEVANT RESIDUES IN SEQ ID NO: amino acids 3334-3344
US-08-986-106-4
Query Match 2.6%; Score 250.5; DB 3; Length 3418;
Best Local Similarity 18.6%; Pred. No. 1.3e-07;
Matches 387; Conservative 280; Mismatches 716; Indels 693; Gaps 96;
QY 85 EFTDFVSKFDFLTNNGRTVLEIPKQYQVVISFSFPE-----DDKERF 126
||:| |::: |::| |:
Db 1126 EPTQF-RKPSYILQ--KSTFEVPENQMTILKTTSECRDADLHVIMNAPSIGQVDSKQF 1182
||:| |::: |::| |:
127 RLGFLHKEKLEDGNTIAQSAKFIYLLPLDMPKAAALQOQSYIVDKN----- 171
||:| |::: |::| |:
1183 EGTVEIKR-----KEAGLLKNDCKNSASG---YLTDENEVGRGFYSAHGTKL 1227
||:| |::: |::| |:
172 -----FNNLIITHPLSNFSAQSIKPLALTRSS--DFIAKLQFNQODELWVYL 216
||:| |::: |::| |:
1228 NVSTALQAKVCLFSD--IENISEETSABVHPISUSSKCHSDSVSMFKIENHNDKTVSE 1285
||:| |::: |::| |:
217 EKFFDLEALKANIRLQTADFSPEKGNLVDPFVYFIRNPQNO----- 258
||:| |::: |::| |:
1286 KNNKCOLLIQNNIEMTT-----GTFVEIETENYKENTENEDNKYTAASRSHNLEFDG 1338
||:| |::: |::| |:
259 -----KEMASDLNQOKTVRLYLRTFSPQAKTILKQYKYKDETFLSSIDLK- 305
||:| |::: |::| |:
1339 SDSKNDTVCIIHKDETDLFTDQHNICLKLSQGFMEKGTQIKE-DLSDLTFLVAKAOE 1397
||:| |::: |::| |:
306 -----ASNGTSLPA--NENDLADQLDVLDDVDFGGQSEITITSNQVKPVPASERSLK 358
||:| |::: |::| |:
1398 ACHGNTSNKEQLTATKTEQNIKD-----FEYSD-----TFFQTASGNISVAKLEFN 1444
||:| |::: |::| |:
359 DRVKFKKQOQKPRIEKFSI-----YEYDALSFY-----SQL 389
||:| |::: |::| |:
1445 KIVNF-FDQKPELHNFSLSLHSDIRKKNKWDILSYEBTDIVKHILKESVPGVGTGNQL 1503
||:| |::: |::| |:
390 QELVSKPNSIKDLVNATL-----ARNRFLSGKYKNFLFD----- 423
||:| |::: |::| |:
1504 VTFQOQPERDEKIKEPTLLGFHTASGKKVKIAKESLDKVKNLFDKEKEQGTSEITFSHOW 1563
||:| |::: |::| |:
424 -----DLASHLDYFLVSKAKIKOSSITKK--LFIPLIKISKSSILGQO 467
||:| |::: |::| |:
1564 AKTLKYREACKOLELACETIETIAPKCKEMQNSLNNDKNLVSIETVPPKLSLNLCKQ 1623
||:| |::: |::| |:
468 EPNITLFEKEVTFKLDNFRDVEIEKAFGLLYPGVNEE-----LEQARKAQRASFEKEKS- 522
||:| |::: |::| |:
1624 TENLKT--SKSIFLKVKHVENVEKETAKSPATCYTNQSPYSVIENSALAFYTSCKRSTV 1681
||:| |::: |::| |:
523 -----KKGLEK--FSQOKEENSKAINNOQGLEEDDNITERLPENSIOQOENAGUG 572
||:| |::: |::| |:
1682 SQTSLLEAKKWLREGIFDQOPERINTA-----DYVGNLYENNNSNTIAEN----- 1727
||:| |::: |::| |:
573 ASPDKPYMIKQVQNRQYVLAKSQIOELIKAKDYTKLAKLSNRHTYNISRLKEQLFDVN 632
||:| |::: |::| |:
1728 -----DRNHL--SEKQDTYLSNSS-----MSNSYSTH-----SDEVYNS 1760
||:| |::: |::| |:
633 PRIPSSRDIEKAKFVLDKTEKNKYWQIYSSAPVFNQKMSLFGYRYLLGLDPKQTIHEL 692
||:| |::: |::| |:
1761 GYLSKNKLDGIEPVLKVEDQK-----NTSPSKVISNVKDAAY-----PQTVNE- 1806
||:| |::: |::| |:
693 VKLGQAGLQPEGYENLPSDFNLEDLKNIRIKTPLPSQKDNPKLSLLDFNNYDGEIKAP 752
||:| |::: |::| |:
1807 -----DICVEEL--VTGSSPCNKNAAIKLSISNSNP---EVGPP 1842
||:| |::: |::| |:
753 ERLGLPLPKELRRNSNGSGSQNSNSPWEQIIISQFKD-----QNLNSODOLA 801
||:| |::: |::| |:
1843 AFRI-----ASGKIRLCS--HETIKKVDIIFDTSFQVKIKENNEKSKIC 1885
||:| |::: |::| |:
802 QFSTKIW-----EKIIGDENEFDQNNRLOYKLLKLDQESWINKTRDNLTYLGD 851
||:| |::: |::| |:
1886 Q--TKIMAGCYEALDDSEDLHNSLDNDECSMHKVFADIQSEELQHNQNM--SGLEK 1941
||:| |::: |::| |:
852 KLVKPKN-NLEAKPRQISNQELLTATYTSNAALNNWNYQDSGAKST----- 899
||:| |::: |::| |:
1942 VSKISPCDVSLETSDICKCSIGKL-----HKSYSANTCGIFSTAGSKSVQVSDASLQNR 1997
||:| |::: |::| |:
900 IIFEEIABLDPKVKKEV-----GADVYQLKPHYAIGFDD---NAGKFNQOEVISSRTI 950
||:| |::: |::| |:
1998 QVFSIEISTQVQFSKVLFKSNEHSDQLTRENTAIRTEPHLISQKGSYVNVNSAFSG 2057
||:| |::: |::| |:
951 YLKTSGSKLEADTIDQLNQAVKNAPLGLQSF-----YLDTERFGVQFKLATSIA 1000
||:| |::: |::| |:
2058 FSTASGK--QVSILESSLHKVGV---LEEFDLIRTEHLSHYSPTSQRNV----- 2102
||:| |::: |::| |:
1001 VQHKQKEKTLPKKLNDNDGYTLIHDKLKKPVIPQISSSPEKQWFEKLNQNGSQNVNVT 1060
||:| |::: |::| |:

Db	2103	-----SKILPRVDKRNPEHCNVANSEMEKTCSEKFKLS-----NNLNVEGGSSENN---2146
Qy	1061	FGSIIESPYSTWFQEDADLDQDQDDSRQGNNSLONQEAAGLLKQKLAILLGNQFIQYYQ 1120
Db	2147	-HSIKVSPYLS-QFOQD-----KQOLVLGTQVSLVENIHLVGKEQ 2184
Qy	1121	QNDKEIBEFILINVEKSELSPRVEFKLAKT-----LEDNGKT-IR 1159
Db	2185	ASPNKVZMEIKTETSDVPKTNIEVCYSYKXSENYPETEAVEIAKAFMEDELDTDSK 2244
Qy	1160	VLSDETMSL-----IVNTTIEK-----TPEMSAVPEVDFKWVEQVDPRTPLAAKTK 1206
Db	2245	LPSHATHSLTCPENEBMWLSNKRIGKRRGEPLILVGEPSIKRNLLNEFDRIENOEKS- 2303
Qy	1207	FVLKFDQOIPVDSGNSIDKWLASIPLVIHQMLRLSPV-----KTIRELGKTEQOQQQ 1262
Db	2304	--LKASKSTP-----DGTIKDR-----RUFMH--VSLEPITCVPTTKE-----2340
Qy	1263	QOQOQOQPKKAVRKEEBLETYNPKDEFNILLPLTKAH--RLTJ-----SNL-VANDPN 1314
Db	2341	--ROEQINPNTA-----PGQEF-----LSKSHLYEHLTEKSSNLAVSGHPF 2382
Qy	1315	YKIBDLVKINEAGDHLAFSLRANNIKRLMNTPIITPADYNPPFYNNEDWRSIDKYLNN- 1373
Db	2383	QOVS-----ATRNEKMRHLITGCRPTKVPFPFKTSHFHRVEQCVRNI 2426
Qy	1374	-----KGNVSSHQQAAGNQGSGLIORLNKNKPE-----TFTPA-----LIAIKOR 1416
Db	2427	NLENRQONIDGHGSDSKNKINDNEIHFQNKNSNQAAAVFTKCEEBPLDI-----2481
Qy	1417	NTNLSNYSD-KITMIKPK-----YLVERSIGVPWSTGLDGYIGSEQTKDGT 1463
Db	2482	--TSQWARDIQMWRIKKQRQRFVPOPGSLYLAKTST-LP-RISLKAAGVGQVPSCSH 2537
Qy	1464	SSSQQKGFQDFQALGKJNTE-----YHGK-----LGLSIRIFPGNELAKIKDASN 1511
Db	2538	QKLYTYGVSRHCIC-INSKNAESQFHTEDYFGKESLWTKGQLADGG---WLIISNDG 2593
Qy	1512	KKGEKLLSYDLFNKVLNEYEKKSPK-IAGKWTNIIHPDOKEYPNPNOKL-----PEN 1563
Db	2594	KAGKEE-----FYRALCOTPGVDPKLSRIWYVNH-----YRWIIWKLAAECAPPK 2641
Qy	1564	YLNILVLPQWKTLYNSDDFITNLVFBEPESDRSGSGTKLKQVTKQVNNYADWGSAYLT 1623
Db	2642	PANRCLS-PERVLLQLKYRYDTEI-----DRSRRSAIKKIMER-----DQTRAKT 2686
Qy	1624	FWYDKNIITNOPNIVITANIADVFIKDVKELDNTKLIAPNITQWPNISGSKEKFFKPTV 1683
Db	2687	VLCVSDIIS-----LSANISSETSKTSS-ADTKQVAIIELTDGW-----2725
Qy	1684	FFGNWENESSMNSQAQPTTWEKIREGFALQALKSSPDQK-----1723
Db	2726	-----YAVKQLDPLPLAVLNG-----RLTWGQKILHGAELVGSPPACTPLEA 2770
Qy	1724	--TRTFVLITNAPLWKYGLPGFQNGFNFKTDQWRLVQFQNDN-----QIAALRVQSDR 1777
Db	2771	PESLWMLKSANSTRPARWYTKLGFPPDRPPPLPSSLFSDBGVGVVDVLIQRAYPIQR 2830
Qy	1778	PEKSS-----EDKQKQKWKFKKVIPEEMP 1802
Db	2831	MEKTSGLYIFRNEREEKEAKTVCAQOQKLEALF 2866

Db	577	---	QTSQNSLPNSTSSNSVLENEKFCGTSIWTAFNAN-1YNLENTKSEYEISTLGNK-	633
Qy	701	LQFEGYENLSDFNLEDLNIRIKTPFOSKDNPKLSLLDFNNYYDGEIKAPEGLPLFL	760	
Db	631	-----LPDFKLVKTNQNL-----ILAQS-----KISL-----NNII-----	658	
Qy	761	PKELRRNSSGGSGONSNSPWEQBIISQFKQO-----NLSNQDLAQFSTKIWEKIG	813	
Db	659	-----NSNKGK-----YDIKKFNPDVFLDGTINYQNQCK-----	688	

RESULT 28
US-09-198-484-2
; Sequence 2, Application US/09198484
; Patent NO. 6162435
; GENERAL INFORMATION:
; APPLICANT: Minion, F. Chris
; APPLICANT: Hsu, Tsungda
; TITLE OF INVENTION: RECOMBINANT MY
; FILE REFERENCE: 19000-028/P028

Qy 1001 VQHKQKXTLIPKLNNDGYTLIHDKLPVPIQISSPEKDMFGKLNQSQSNVNVST 1060
Db 1822 -----SKLPRVDRKNPEHCVCNSEMERTCKSEFKLS-----NNLNVEGSSENN--- 1865
Qy 1061 FGSIIESPYFTNFQEDADLDQDQDSRQGNNSLDNQEAGLLKQKLAIIIGNOFIOYYQ 1120
Db 1866 -HSIKVSPYLS-QFQD-----KQQLVLGTQKSVLNIHVILGKEQ 1903
Qy 1121 QNDKEIEFEIINVRKVSLSRPFVKLAKT-----LEDNGKT-IR 1159
Db 1904 ASPKNVMEIGKTTETFSVPVKTNIIEVCSYKSDSENFEYFEAVEIAIAKAFMEDDELTDK 1963
Qy 1160 VLSDETWSL-----IVNTIIEK-----TPENSAVPEVDTKWQYDPTPLAAATK 1206
Db 1964 LPSHATHSLFTCPENEEWLSNSRIGKRGREPLILVGEPSIKRMLNLEFDRIIENQES- 2022
Qy 1207 FVLKPKDQIPVDGSGNISDKWLASIPLVIHQMLRLSPV-----KTIRELGLKTEQOQQQ 1262
Db 2023 --LKASKSTP---DGTIKDR-----RLFVHH--VSLEPIITCVPRFTTKE----- 2059
Qy 1263 QOQOQOQOQOQKAVRKEBELETYPNPKDEFNINPLTKAH-----RLTL-----SNL-VNNDPN 1314
Db 2060 --RQBIQNPETA-----PGQEF-----LSKSHLYEHLTLEKSSSNLAVSGHPF 2101
Qy 1315 YKIEDLVKINEAGDHQAFSLRANNIKRLMNTPIITFADYNPFYVYNEDWRSIDKLYNN- 1373
Db 2102 YQVSG-----NKNG-----KMRKLITGTPTKVFPPFKTHFRVQCVNRI 2145
Qy 1374 -----KGNVSSHQQOAGGNOGSGLIQRLNKNIKPE-----TPTPA-----LIALKDR 1416
Db 2146 NLEGNRQKQINDCHGSDSKINDNEIHQFNKNNSNOAAVTFKCEEPDLDI----- 2200
Qy 1417 NNTNLNSYD-KIIMIKPK-----YLVRSIGVPMSTGLDGIYIGSEOTKOTS 1463
Db 2201 --TSLQARDIQDMRIKKQRQRFVFPQGSILYAKTST-LP-RISLKAAGVQVPSACSH 2256
Qy 1464 SSSQKQGFDDFIQALGLKNTE-----YHGK-----LGLSIRIFDPGNELAKIKDASN 1511
Db 2257 KQLYTYGVSKHCIR-INSKNAESFOFHTEDYFGKESLWTGKGIQADGG---WLIPSDNG 2312
Qy 1512 KKEGKLLKS 1521
Db 2313 KAGKEEFYA 2322

RESULT 30
US-09-150-867-1
Sequence 1, Application US/09150867
Patent No. 6645748
GENERAL INFORMATION:
APPLICANT: Wood, Kenneth W.
APPLICANT: Sakowicz, Roman
APPLICANT: Goldstein, Lawrence S.B.
APPLICANT: Cleveland, Don W.
APPLICANT: The Regents of the University of California
TITLE OF INVENTION: Plus End-directed Microtubule Motor Required for
FILE OF INVENTION: Chromosome Congression
FILE REFERENCE: 18557C-00011005
CURRENT APPLICATION NUMBER: US/09/150,867
CURRENT FILING DATE: 1998-09-10
EARLIER APPLICATION NUMBER: US 60/058,645
EARLIER FILING DATE: 1997-09-11
NUMBER OF SEQ ID NOS: 11
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 1
LENGTH: 2954
TYPE: PRT
ORGANISM: Xenopus sp.
FEATURE:
OTHER INFORMATION: Xenopus centromere-associated protein-E (XCENP-E)
OTHER INFORMATION: member of the kinesin superfamily of microtubule
OTHER INFORMATION: motor proteins
FEATURE:

; NAME/KEY: DOMAIN
; LOCATION: (1)..(472)
; OTHER INFORMATION: kinesin like motor domain
; FEATURE:
; NAME/KEY: DOMAIN
; LOCATION: (473)..(2752)
; OTHER INFORMATION: rod domain
; FEATURE:
; NAME/KEY: DOMAIN
; LOCATION: (2753)..(2954)
; OTHER INFORMATION: tail domain
US-09-150-867-1
Query Match 2.5%; Score 242.5; DB 4; Length 2954;
Best Local Similarity 17.7%; Pred. No. 3.6e-07;
Matches 376; Conservative 346; Mismatches 670; Indels 727; Gaps 95;
Qy 28 LASKVKYRGVNPQTQGVISQGLI-DSVAFKPSI-----ANFTSDY 66
Db 496 LASKVTHREKTSLHQSMIDFQISDSVQFHDSSKENQLYLPKDSGDMACRKAPEKEI 555
Qy 67 QSVKALLNG-----KTFDPKSEFTDPVS-----KFDPLTN-----NGRTVLRIK 108
Db 556 TSLQOQLQSKKEEKELVQSFELKIAELEBQLSVAKNLEWVNSRHSINAEVQTDVEK 615
Qy 109 KYQVWISFSPEDDKERFLGFLHKEKLEDG----- 139
Db 616 --EVVRKEMSVLGDGSGYNASNDLQDSSVDGKRLSSSHDCEIHRKMLEQKIVDLEEFIE 673
Qy 140 -----NIAQSAATKTYLLPLDMPKAAALQYSYIVDK-----NFNLIHPLSNFSAQ 186
Db 674 NLNKKSENDKQKSEQDFMESIQLECAIMAEKANALBELALMRDNFNI-----LENETLK 730
Qy 187 SIKPLA-LTSSDDIAKLNQENQDELWVLEKFFDLEALKANIRLOTADSFEGNLVD 245
Db 731 --REIADLERS-----LKENQETNEFEI---LEKETQKE-----HEAQLIHIGSLK- 772
Qy 246 PFVYSFRNPNQOKEWASDLNODOKTVRLYLRTFSPQAKTILKDYKYKDETF---LSSI 302
Db 773 -----KLVENAEYNNQNLLEEDLET-----KTLLKKEQEIQLAELKRADNLQKVRNF 820
Qy 303 DLKASNGTS-----LFANENDLKD-----QLDVLDDVSDYFGG 336
Db 821 DLSVSMGDSEKLCBEIFQLKQSLSDAEAVTRDAQKCEGSLFASSENLEKEMEDTSNWT-- 878
Qy 337 QSETITSNVQKVPVPSERSLK-DRVFKK---DOQKPRIEKFSLYEY-DALSFYSQLOE 391
Db 879 -----NQKKAASLFEKQLETEKSNYKKMEADLQKELQSAFNEINYLLAGKVPKD 931
Qy 392 LVSKPNSIKDLVATLARNLRFSLGKYNFLPDDLASHLDYFVLVSKAKIKOSSITKGLFI 451
Db 932 LLSRVELEKVV--SEPSKQLEKALEEKNALENEVTCLESEKFLPNEVECLKNQISKASEE 989
Qy 452 ELPIK-----ISLKSSILGDOEPNIKTLFKEVTFKLDNFRDVEIEKAFGLLYPGV 502
Db 990 IMLLKQEGEHSASIIISKQEIIMQEQSEIQLQTD-EVTHTOSKVOQTE----- 1036
Qy 503 NEELEQAKQARASFEKESKGLKFSQOKEENSKA-----INNQEGLBEDDNI 553
Db 1037 -EOYLEMKKMHDDLFKEKYIRNKSEADLLRWEMLKGTWESVEVKIADTKHELEETIDK 1095
Qy 554 ERLPENSPYOQENAGLSPDKPYMIKDVQN-----ORYYLAKSQIQELIKA 602
Db 1096 EQLH-----EKYFFQAMQTFPIITPLSDSLPPSKLVEGNSQDPIEI 1138
Qy 603 KDYTKLAKLNRHTYNTISLR-----LKEQLFDVNPRIPS---SRDIEKAKFVLDTEKNK 655
Db 1139 NDYHNLIALATERNNIMVCLTERNSLKEQVIDLNTQLQSLQAQSIKESDLQKPKQDLEE 1198
Qy 656 WQIYSSASPVFQNKWSLFGYRYRLLGLDPKQTHIELVK---LGOKAGIQEGYENLPSD 712
Db 1199 -----GEVKLLLE-----ELDKGHJLTDLSQSLIEKLEULENLEVT 1232

Qy 381 DALSFYQLOELVSKPNSIKDLVNATLARNLRS-----LGKYNFLPDDLASHLDYYFLV 435
Db 2081 SG-----NVSSPKH-----AMMRFTEDRLDRGREKLIYEDVRD----- 2115
Qy 436 SKAKIKOSSITKLFIELPIKISLSSILGQEBNITLFEKEVTFKLDNFRDVEIEKAF 495
Db 2116 --TVKEA-----EKLTEVSQFFRD-KTEK-- 2137
Qy 496 GLLYPGVNLEQARAKQASFEKEKKGKGLKFSQOKEBNS-----KAINNQEGLEEDNI 552
Db 2138 -----LNDELQSPK-----KARPNG-KYSSQSPTSSSEKVLTLTLLASNDWV 2183
Qy 553 TERL--PENSPIQOQENA-GLGASDPKPYMIKDQORYYL-AKSOIELIKAKD----- 604
Db 2184 KARQHGPDGQGFPAEAKAPSLSPSEKMWLSQTESKSTVEAKGSLQ-SKAPDGPQS 2242
Qy 605 --YTKLAKLLNR-----HTYNISRLKQLFDVNPRISSRDIEKAKFVLDKTEKNK 655
Db 2243 GFQLKQSKLSIRLKFEGTHAKSKDMQSDRKGSGQSRIPVK----- 2286
Qy 656 YWQIYSSASPVFNKMSLFGYRYLLGLD-PKQTI--HELVKLGQAGLOFGEYENLPS 711
Db 2287 ---IQESKLPVYQ---VFAREKQQAIDLDPDESVSQKDPMWLTK-DEHAQNEIIVN 2338
Qy 712 DPNLEDKNIIRIKTPLFSQKDNFKLSLLDFNNYDGEIKAPFGLPLPKELRNSNS 771
Db 2339 DSGSDNVKQKTEMSSKAMPDSF-----SEQOAXD--LACHITSDLA----- 2378
Qy 772 GGSQNSNPWEOELISQPKDQNLNQDLAQFSTKIWEKIIGDNEFDQNNRLOVKLLKD 831
Db 2379 -----TRGFWKXVF-----RTWES-SGATNNKSQEKLSHVLVHD 2413
Qy 832 LQESWINKTRDNLWYTLGDKLVKPKNNLBAKFRQISNLQELITAFVTSAAALNNNNY 891
Db 2414 VRENHIGP-----ESKSVQDN-----EFMSVTEREKL-----LTNGSL----- 2450
Qy 892 QDSGAKSTIIIEETAEIDPKVEKGVADYQKHYAIGFDNAGKFNQEVIRSSRTIY 951
Db 2451 -----ETKEMTVKSPK-----KVLRYEYVYKGDHPGGLDQPSRRS----- 2488
Qy 952 LKTSKSKLEADTIDOLNAQVKNAPLGL--QSFVLDTERFVGFKLATSLAVQHKQKEKT 1009
Db 2489 -ESSAVSHIPRVADERRMLSNIPDGFCEQASAPKHE-----LSQKLSQSSMSKETV 2540
Qy 1010 LPKKLN--DGYTLIHDKLK-----PVPIQISSPEK-----DWFEGKLNQNGQ 1052
Db 2541 ETQHENSIEDEKVTYSEISKVSKHQSVGLCPLEETETSPKSPDSLEFSFGK----- 2594
Qy 1053 SONVNVSTFGSIIESPSTFNPQEDADLDQDQDSDRQNNSLDNOEAGLLKQKLAILLG 1112
Db 2595 -----ESP-----SSDVFDHSPIDGLEKLAPLAQTEGGEIKTLPVVY- 2632
Qy 1113 NOFIOYQONKDEIEFEIINVEKVELSFRVEFKLAKTLEDNGKTRIVLSDETMSLIYNT 1172
Db 2633 -SFVQVGQYKEIQ-----QGVKKIISOECKIVQ-----ETRGTFYTT 2671
Qy 1173 TIEK-----TPEMSAPVEV--PDKWVEQYDPTPLAAKTKFVLKFDQIPVDGSGNI 1223
Db 2672 RQKQPPSPQSGPEDTLTEQVSFLDSGKSLPTPETPSSEVSY--EFTSKTP----- 2722
Qy 1224 SDKWLASTPLVTHQKRLSLPVVKTIIRGLKTEQOQQOQQOQQOQQOQKAVRKEELE 1283
Db 2723 -DSLAIYIP-----GKPSPIPEV-----SESEEEEOAKSTSLKQTTVEETAIVE 2765
Qy 1284 TYNPKDEFNINPLTKAHLRTL-----SNLVNND-----PNYKIEDLVK-IKNE 1326
Db 2766 REMPNDVSKDQNRKPNKRVAVIEPPPPPLDADQIESDKGHYUPEKEVDMEIENVLODE 2825
Qy 1327 AGDHQLAFSLRANNIKRLMNTPI--TFADYN-----PFFYINEDWRSIDKYLNN 1373
Db 2826 HDKYQLAEPV-----IRVQPPSPVPCADVSDSDDESIVQPVVKYTFKLKEVDDEQKE 2881
Qy 1374 KGNVS-----SHQOQAGNGQS-----GLIQRILNKIKPFTTALLALKDRNNTNLSN 1423

Db 2882 KPRAAEKASQNELESNGSGKDNERGLGLDSPQNE-----IAQNGNNDQSITE 2930
Qy 1424 YSKIMIKPKYILVERSIGVPMWSTGLDGYIGSQTKDGTSSSSSQ----- 1468
Db 2931 CS-----IATTAFFSHDTDATEDSLDGY-DLQDEDDGLTESDKLPIQAMEIKKOIWT 2984
Qy 1469 -----KGFDDQFIQALGIKNTEYHGKGLGLSIRIPDGNELAKIKDASNKGEELKLSYDL 1524
Db 2985 EGILKPADRSFQS-KLEVIEEGKV-----PDEKPPSKSSSEKTPDKT----- 3030
Qy 1525 FKNYLEYKSPKIAKGTNIHPDQKEYPNP--NOKLPENYLNVLNQPMKVTLYNSSD 1582
Db 3031 -----DQKSGAQPFLLSRRHPRDSRVFDDTVFSYKVDSEFAT-----PEKTVATKGLD 3077
Qy 1583 FITNLVFBEPGSRG----- 1597
Db 3078 F-----DPMNNRNGDEVDPSKSRDETKPFGFLAVEDRSRPAITPTTPARTPDESPTT 3131
Qy 1598 -----SGTKLQVIOQVANNYADWGSAVLTFWY-----DKNIIT 1632
Db 3132 SEPNPPFHEGKPFEMTRSGAIDMSKRDFFVEERLQFFQIGHEHTSEKSGQGGEDKSMVT 3191
Qy 1633 NQPNVITANIADVFIKDVKELEDNTKLIAPNIQWENI-----SGSKERPFYK 1680
Db 3192 ATQO---PQSGDITVEINLERNVETPTVEPN---PSIPTSGECQSGTSSSGSLEKSA 3243
Qy 1681 PTVFFGWNWENSS-----MNSQATPTWEKIREG---FALQALKSS---FDQK 1723
Db 3244 AT-----NTSKVDPKLRTPIKMGISASBTMTMKKEGCEITDKIEAVMTSCQGLENE 3294
Qy 1724 TRIFVL--TTNAPLPLWKYGPLGQNGENFKTQDWRLVFQNDNQIALRVQ----- 1773
Db 3295 TITMISNTANSQMGVRPHEKHDFQK-DNF-----NNNNLSDSTIQTDMNIMNIV 3343
Qy 1774 -----EODRPEKSSSEDK-----DKOK 1789
Db 3344 LTHESAPTCTTERKDNFVKVSSGKKTGVLQGHCVDRKQK 3381

RESULT 32

US-09-949-016-6978
; Sequence 6978. Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6978
; LENGTH: 4377
; TYPE: PRT
; ORGANISM: Human
; US-09-949-016-6978

Query Match 2.5%; Score 238.5; DB 4; Length 4377;
Best Local Similarity 17.9%; Pred. No. 1.2e-06;
Matches 373; Conservative 286; Mismatches 690; Indels 729; Gaps 100;
Qy 35 RGVNTPQVISOQLGLISVAFKPSIANFTSDYQSVKALLNGKTF---DPKSSSEFTDFVS 91
Db 2174 RSYDPSAGDVPTQTPERPSPKPS-----PTFMELEPKPTT----- 2209


```

; CURRENT APPLICATION NUMBER: US/09/543,681A
; CURRENT FILING DATE: 2000-04-05
; PRIOR APPLICATION NUMBER: US 60/128,706
; PRIOR FILING DATE: 1999-04-09
; NUMBER OF SEQ ID NOS: 8344
; SEQ ID NO 4980
; LENGTH: 2807
; TYPE: PRT
; ORGANISM: Proteus mirabilis
; US-09-543-681A-4980

```

Query Match	2.4%	Score 233;	DB 4;	Length 2807;	
Best Local Similarity	20.0%;	Pred. No. 1.5e-06;			
Matches 329;	Conservative 227;	Mismatches 572;	Indels 518;	Gaps 90;	
Qy	503	NEE-LEGARKA-ORASPE---	KEKSKGLKEFSQKEENSKAINNOGLEEDDNI	TERLP 557	
Db	14	NQELDDAQQAIRNLSINAGIRKSANVLPSTSD---		---LNGDDVATYQCP 60	
Qy	558	--ENSPIQOQENA-----	GLGASDPKPYMKDVQNRYYLAKSOIELIKAKVYTKLA	609	
Db	61	SLRSGNLQYHKIGAIANVLSGLGESP-----		---IFETLKQKFNVK-- 98	
Qy	610	KLJLSNRHTY-----	NISLRKQOLF--DYN-----	---PRIPSSRDIE 642	
Db	99	--IKNRNTYPTFILEKNSGRIOKNIYLDNINDPTQOKII	EHFILANYSNFAMPENLHLE	156	
Qy	643	KAKFVLDETKKWTQIYSSASPVQNKWSLFGVYRVLGLD	PK-QTIHELKVLGQ-KAG	700	
Db	157	NDKTKSSHNNKDRILABY-----	VEHKWQKL-----	---TDSKWSVDELHKVASVKGK 202	
Qy	701	LOPEGVENL-----	PSDENLEDLKNIRIKTLPFSQDNFKLSLD---	---740	
Db	203	IQGESYKSIITLSEYHEKESNTNCLPIEKRLVLAQLR	NOIDGYLLCHPDSERNDGLKQL	262	
Qy	741	-----	FNNYVDGEIKAPFGLPLFLPKELRN---	---767	
Db	263	QTQVNTRYKHASALIQDSHTINADSSFSRLYDLSNA---		---NLKRSKHIVIDI 311	
Qy	768	-----	SNSGGSGNSNPWEQIITSQFKDQNLNQDQLAQFSYKI	WEKILGDENEFD 819	
Db	312	NGHFVTEGKNNLQGNRLTS--GKEVIERVK--	NAVEKEYCAPIANKVFQSQA--	---NNEFA 365	
Qy	820	QNNR-LQYKLLKDQLQ--	ESWINKTRDNLX-WTVLGDKLKVKPKN--	---NLEAKFRJOJSLNQ 872	
Db	366	ENGEIDVSGLKKTHRAIERELSPHSATLYIW-----		---KPSNHSNLGHATLQIQGR 416	
Qy	873	ELLTAFYTSAAJLN--	NWNYQDSGAKSTIIPBEIAELDPKVKEKVGADV-----	---YQ 922	
Db	417	LLLN--	TEDQNNHKNYISWPNKAKS---	YEDVMGVPSQPKESLASDVAABENDGFG 470	
Qy	923	LKFHYAIGFDNAGKNQEVIRSSRTIYLKTSKSKLEAD	ITDQLNQAQVNA--	---PLGLQ 980	
Db	471	LNKYK-----	KELNRYDLRMWIAKETV-TTADDSLADAKNULTIPNYI	ENSKIPIRTIG 524	
Qy	981	SPYLDTERFGVQK-----	LATSL--AVOHKOKEKTLPKLLNNDGYTLIHK--	---LKKVP 1032	
Db	525	QPFID--	EWVITQKLNANIATNFLAEVQKFSGNTI--	---TINGAMILLTPTLLSKRKIP 579	
Qy	1033	QISSSPKDWPECKLNQNGSQNVNVSFTGSIIESPYSTN	TFDADLD-----	---QD 1083	
Db	580	LSVAQPFTEQWD-----	NIKT-----	---QQSDISIQLLKAMTWIQE 614	
Qy	1084	GODDSRQGNNSLDNQEAGLLKQKLAILLGNQFIQVYQ	ONDKEIFEI-----	---INVEK 1135	
Db	615	GKDP-----	LTKKRIA-LVENEFAEQKSIDRPKTSLPDEGRV	FRINLEG 660	
Qy	1136	VSELSFRVEFKLAKTLENGKTRIVLSDETMSLIVNTT	IEKTPEMSAVPEVFDTKWVEQY	1195	
Db	661	LDVAAMQAEWORIKN-SPNGR-YHLLTQNCSSIVA---		---KVLKAGGADELIGYNWQPKY 713	
Qy	1196	DRTP-----	LAAKTKFVLKFKDQ--	IP-----	VDGSGNISDKWLASIPLVTHQOMRL 1242

```
; LENGTH: 3210
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (0)..(0)
; OTHER INFORMATION: Polypeptide Accession Number P49454
US-09-538-092-1154

Query Match      2.4%; Score 231; DB 4; Length 3210;
Best Local Similarity 18.1%; Pred. No. 2.5e-06;
Matches 371; Conservative 364; Mismatches 783; Indels 534; Gaps 92;

Qy 61 NFTSDYQSVKALLNGKTFDPKSEFTDFVSKDFDL-TNNGRTVLEIPKXYQVVISFSP 119
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1066 NRKNELEQLKEAF-----AKEHQEFLTKLAFABERNQNLMLELTVQALRSEMTD 1116

Qy 120 EDDKERFRLGHLKEKLEDGNIAGSATKFIYLLPLDMPKAAALQYYSIVDKNFNLI--- 176
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1117 NON-----NSKSEAGGLKQE-----IMTLKEQNK-----MQKEYNDLLQEN 1153

Qy 177 -----IHLPSNFAQIKPLALTRSSDFTIAKLNQ--FNNQDELWV-----YLEK 218
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1154 EQLMKVWKTHCEONLESEFIRNSVKERESE-----RNQCNFQMDLVEKIEISLDSYNAQ 1209

Qy 219 FFDLEALKAN-----IRLQTADPSFEKGNLVP-----PFV 248
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1210 LVQLEAMLRNKLQLOSEKEKECLOHELQTIIRGDLTSLNLDWQSOEISGLKDCEIDAE 1269

Qy 249 YSFIRNPONKEWASLDNQOKTVRLYLRTFS-----POAKTILKDYKYKDETFSSID 303
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1270 EKIYSGPHE-----LSTQNDNAHLQCSLQTTMKNLNELEKICILQAEKYELVTELNDSR 1325

Qy 304 LKASNGTSLPANE-----NDLKQDLDVLLDSDYFGQ-SETITNSQVKKVP 351
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1326 SECITATRKAAEEVGKLLNEVKILNDSDGLLHSEL--VEDIPGSRGEQNPQHPVSLAP 1383

Qy 352 ASERSLKORVKFKDQOKPRI-----EKFSLYEYDALSFYSQQLVSKNPSIKDLVNATL 407
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1384 LDESNSYEHULTSDKEVQMHFAELQEKFLQSEHKILHDHQMSKMSSELQTYVDLSK 1443

Qy 408 ARNLRFSLGKYNF---LFDLLASHLDYFL--VSKAKIKOSSITKL-----FIELP 454
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1444 AENLVSTNRNFGDLVKEMQGLGEBGLVPSLSSCCVPSSSLSSLGSSSFYRALLEQT 1503

Qy 455 IKISLSSSILGDQEPN---IKTLFEKEVTFKLNFRDVEIEKAAGL-----LYPG 501
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1504 GDMSSLNLSGAVSANQCSVDEFCSSLQTYVDSLKAENLVLTNLNFGDGLVKEMQLG 1563

Qy 502 VNBELBOARKAQRASFKEKSKGLKEFSQ---OKGENSKAINNOBLE-----547
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1564 LEBGLVPSLSSSCVPSSSLSSLGSSSFYRALLEQTGDMSSLNLSGAVSANQCSVDEVF 1623

Qy 548 ----EDDNITRLPENSPIQYOENAGLGPSPKPYMIKDVONQRYYLAKSOIELIKAK 603
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1624 CSSLOEENLRKTFPSAPAGVBEBSL-----CEVYR----QSLKLEEKWESQGIMQNK 1675

Qy 604 DYTFLAKLNSNRHTYNISLKLKQLFDVNPRIPISSRDIEKAKFVLDKTEKQKYWQIYSSA 663
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1676 EIOLEQLLSERQELDCLR-KQYL-----SENEQHQKLTSTV 1712

Qy 664 SPVFQNKWSLFGYVRYLLGLDPKQTHLVLKQKAGLQPEGYENLPSPFNLEDLKNRI 723
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1713 TLEMESK-----LAAEKKQTEQLSLEL-EVARLQIQGL-----DLSRSLLIGDIT 1756

Qy 724 KTLFSGKONFKLSLLDFNNYDGEI--KAPFGLPLFLPKELRN-----767
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1757 EDIAIQGNES-----CDISKEHTSETTERPKHDVHOICDKDAQQDLNLDIEKITGTAV 1811

Qy 768 --SSNCGGSQNSPWE-----QETISQ--FKDQV-----LSNQDOLAQFSTK 806
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1812 KPTGECSGEQSPDTNYPEPCGEDTKQSSSECISLSFSGPNALVPMDFLGNQEDIHNLQLR 1871
```


Db 1914 ERIOEKTDIDRLQELLCASNLQLEBAEQOIQEERELLGRQKEMKAERGVP--E 1970
Qy 1650 VKELEDNTKLIAPNITOWPNISSGKFKYPTVFFGNWENENSSMNSQAQTPTWKIRE 1709
Db 1971 QOLLQETEKLM-----KEK-----LEVQQA---EKVRD 1996
Qy 1710 GFALQ--ALKSSFPQKTRTFVLITNAPLPLWKYGPLFGQNGPKTQDWRLVFNQDNDQI 1767
Db 1997 DLQKQVKALEIDVEQVSRFI-----ELEQKNTLMDLRQONQALEKQL 2041
Qy 1768 AALR--VOEQ--DRP-EKSSSEKDKQKWKFKVVP 1798
Db 2042 EKRRKFLDEQAIDREHDFVQEIQLKLEQOLKVPP 2077

RESULT 36
US-09-538-092-1312
; Sequence 1312, Application US/09538092
; Patent No. 6753314
; GENERAL INFORMATION:
; APPLICANT: Mansfield, Traci A.
; TITLE OF INVENTION: Protein-Protein Complexes and Method of Using Same
; FILE REFERENCE: 15966-542
; CURRENT APPLICATION NUMBER: US/09/538,092
; CURRENT FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: 60/127,352
; PRIOR FILING DATE: 1999-04-01
; PRIOR APPLICATION NUMBER: 60/178,965
; PRIOR FILING DATE: 2000-02-01
; NUMBER OF SEQ ID NOS: 1387
; SOFTWARE: CuraPatSeqFormatter Version 0.9
; SEQ ID NO 1312
; LENGTH: 2472
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (0)...(0)
; OTHER INFORMATION: Polypeptide Accession Number Q13813
US-09-538-092-1312

Query Match 2.4%; Score 229.5; DB 4; Length 2472;
Best Local Similarity 19.0%; Pred. No. 2.2e-06;
Matches 377; Conservative 287; Mismatches 618; Indels 699; Gaps 107;

Qy 182 NFSAQSIKPLALTRSSDFIAKLNQFN-----QDELWVLEKFFDLEALKANIRLQ 232
Db 515 SLSAQEEK---ITALDEFATKLIQNHHYAMEDVATRDAL---LSRNALHERAMRRRAQ 568
Qy 233 TADFSPEKGNLVPFVYSFIRNPONQKEMASDLNQDQKTVRLYLRTFSPQAKTILKDYK 292
Db 569 LAD-SFH-----LQQFPRDSDELKSW--VNEKMKTA-----TDEA 600
Qy 293 YKDETFL-SSIDLKASNGTSLPANEN--DLKDQLDVLDDVDSYFGQSETITNSQV-- 347
Db 601 YKDPNSLQGVQKHQAPAEALSANQSRIDALRAGAKQLIDVNHY--AKDEVAARMNEVIS 658
Qy 348 ---KPPVASE-RSLKDRVKFKDQKQPKIEKFSLEYVDALS FYSQQLVELSKPNSIKDLV 403
Db 659 LWKLLJLEATELKGILKEANQQOQFNRNVEDIELWLY-----VEGHLSDDYKQDUT 711
Qy 404 NATLARNLRFSLGKYNFLFDDLASHLDY-----YFLVSKAKIKQSSITKK 448
Db 712 N---VQNLQ---KHALLEADVAHQDRIDGVTIQAQFQDAGHFDPAENIKKKQEAALVAR 765
Qy 449 LFIELPIKISLSILGDQEPNKTITFEKEVTKLNFDFRDNVELEKAF----- 495
Db 766 Y-----EALKEPVMARKQKLAADSLRLOQL-----PRDVEDEETWIREKEPIAASTNR 812
Qy 496 GLLYPGVNEBLE--QARKAQARAFE---KEKSKKGLKFP--SQKEENS KAINNQGLEEDD 550
Db 813 GKDLIGVQNLKKHQALQAEIAGHEPRIKAVTQKG-----NAMV---EEGHFAAE 859

Qy 551 NITERLPENSIQYOQENAGLGPDPKPYMIKDVQNRYYL-----AKSQIQE---LIKAK 603
Db 860 DVKAKLHE---LNQKWEALKAKASQRRQDLEDLSLOAQYFADANEAEASWREKEPIVGST 916
Qy 604 DYTEK-----LAK---LLSNRHTYINISLR-LKEQ-----LF 629
Db 917 DYKGDSDSAEALLKHEALMSDLSAYSSSIQALREQAQSCROQVAPTDDDETGKELVLALY 976
Qy 630 DVNPRIPSSRDIEKAKFVLDKTEKNKYWQIYSSASVPFNQKWSL-----FGYRYVLLG 682
Db 977 DYQEKSPREVTWKKGDIILLNSTKDW-----WKVEVNDRQGFVPAAYVKK 1023
Qy 683 LDPKQT-----IHELVLKLGQK- 698
Db 1024 LDPAQASRENLEEQGSIALRQEQIDNQTTRITKEAGSVSLRMKQVVEELVHSLLEGEKR 1083
Qy 699 AGL-----QFEGYEN-----LPSDFNLEDLKNIRIKTPLFS--OKDNFKL 736
Db 1084 KGMLEKSCCKFPLFREANELQWINEKEAALTSEEVGADLEQVEVLQKKFDDFQKD-LKA 1142
Qy 737 S---LILDFNNYYDG-----EIKAPE---FGLPLPLPKELRRNSNSGGSQNSNP 780
Db 1143 NESRLKDINKVAEDLESEGLMAEEVOAQVQQEVYGM-----MPREDTDS---KTASP 1191
Qy 781 WEQ-----EIIQFKQNLNSQDQLAQF-----STKIW--- 808
Db 1192 WKSARLMVHTVATFNSIKELNERWRSLOQLABERSQJLGSASEVQRFHRDADETKEWIEE 1251
Qy 809 -----EKITGDENE?---DONNRL-----QYKLLKDLO 833
Db 1252 KNQALNTDNYGHDLASVQALQRKHGFERDLAALGDKVNSLGETABERLTQSHPESAEDLQ 1311
Qy 834 ESWINKTRDNLWYTYLGDKLKVKPKNNLEAKFRQISNLQELLTAFTVTSAAALSNNWYYQD 893
Db 1312 EKC---TELNQAWSSLG-----KRADQKAKLGDSDHDLRFLSDFRDLMSWIN----- 1356
Qy 894 SGAKSTIIFEEIAELDPKVKKEGVADVQLRPH---YAIGFDDNAG-----KFNQEVIRS 945
Db 1357 -GIRGLVSSDELA-----KDVTAEGAL-LERHQEHRTEIDARAGTFFQAFEGFQGOALLAH 1408
Qy 946 SRTYLYTKSGSKLEADTIDQLNQAQVA-----PLGLQSF----- 982
Db 1409 GH---YASPEIKOKL--DILDQERADLEKAWQVRMMMLDQCLEQLQFHRDCEQAENWMAA 1463
Qy 983 ---YLDTERFGVFKLATS LAVQHK-----QKEK-----TLPKKNNDGVTLTHD-- 1024
Db 1464 REAFLNTEKGDSDLSVEALIKKHEDFDKAINVQBEKTAALQAFADQLIAAGHYAKGDIS 1523
Qy 1025 -----KLKXPVPIQISSSPEKD-----WFEGKLNQNGOSQNVNVS 1059
Db 1524 SRNREVLDNRWRRLKQAMIEKRSKLGESQTLQFSDRDVDEIEAWISEKL-QTASDESYPK 1582
Qy 1060 TFGSIIIESPY-----FSTNFOEDADLDQDQDSDSQGNNSLD-----NOEAGLLKOKLA 1108
Db 1583 T---NIQSKHQKHQAPAEALHANADRIKGV1D---MGNSLIERGACAGNEDA--VKARLA 1634
Qy 1109 ILLGN-QFI-----QYQQNDKEIBEIINVEKVSLSF---RVEFKLAKTLEDNGK 1156
Db 1635 ALADQWQFLVQKSAEKSQKLKANKQNFN-----TGIKDIAFWLSEVEALLAS--EDYCK 1688
Qy 1157 TIRVLSD-----ETWSLIVNTTITKTPMSA-----VPEVFDTKWVE-----QYDP 1197
Db 1689 DLASVNNLLKKHQLLEADISAHEDRLKDLNSQADSLMTSSAFDTSQVKDKRDTINGRFQK 1748
Qy 1198 RPLAAKTYFVLKFKDQIIPVDSGNSISDKWLASIPLVIHQOMRLSPVVKTTRELGLKTE 1257
Db 1749 IKMSMAASRRAKL-----NESHK-----LHQFFRDMDDDEESWIKKLLVG 1788
Qy 1258 QOQOQOQOQOQOQKQKAVRKEEELETYNPKDEFNILNPLTKAHLRTLSNLVNNNDPNYK- 1316
Db 1789 SEDYGRDLTGVLNLRKKHKLLEAELAAHEPA-----IQGVLDTGKKLSDDTTIGKEIQOR 1844

QY 1317 ----IEDLVKINAGDHOLAFSLRANNIKLMMNTPTTFADYNPFYFNNEDWRSIDKVLN 1372
DB 1845 LAQFVHVKELK-----OLA-AARGORLESL-----EYQOFVANVEE---EEAWIN 1887
QY 1373 NKGNVSSHQQAAGNOCGSL--IQLMKNIKPETFTPALIALKDRNNNTLSNYSDKIIM 1430
DB 1888 EKWTL-----VASEDYDGTAAIOGLLK--KHEAFETAFTVHKDRVNDVCTNGQD---L 1936
QY 1431 IKPKYLVERSIGVPMSTGLDGYI-----GSBQTKDGTSSSQ--QKGFQDQFIQALGLKN 1483
DB 1937 IKKNHHEENISSKMK-GLNGKVSLEKAAQKANKVNDENSAFLQFNWKADWVES----- 1990
QY 1484 TEYHGKGLSIRIPDPGNELAKIKDASNKGBEKLKSYDILFKNVLYEYKSPKIAGW 1543
DB 1991 --WIGEKENSUKTDDYGRDLSSV-----QTLTKQETFDAGLQAFQO-----EGI 2033
QY 1544 TNI-----HPDQKEYPNPNQKLPENYLVNLQPNKVTLYLNSSSDFITNLVPEPE 1592
DB 2034 ANITALKDQLAAKHVQSKAIEARHASLWKWSQLLAN-----2071
QY 1593 GSDRGSGTKLKQVIOKQVNNNVADWGSAYLTF-----WYDKNIITNQPNVITANIA 1643
DB 2072 -----SAARKKKLEAQ--SHPRKVEDLFTPAKKASAFNWF-NABEEDLTPVRCNSL 2123
QY 1644 DVFIKDVKELED-----NTKLIAPNITQW-----WPN1 1671
DB 2124 EE-IKALREAHDAFRSSLSAQAQPNQLAELDRQIKSPRVASNPYTWTFMEALBETWRNL 2182
QY 1672 SGSKEKFKYPTVFFGNWENENSMNSQAQTPTWKIREGFALQALKSSFD---QKTRTFV 1728
DB 2183 ----QKLIK-----BRELEQKQRRQEBENDKLQOEFQAHA--NAFHQWQIETRYL 2228
QY 1729 L 1729
DB 2229 L 2229

RESULT 37

US-08-353-700-1
; Sequence 1, Application US/08353700
; Patent No. 5599919
; GENERAL INFORMATION:
; APPLICANT: YEN, TIMOTHY J.
; APPLICANT: RATNER, JEROME B.
; TITLE OF INVENTION: NUCLEIC ACID ENCODING A
; TRANSCRIPTION: TRANSCRIPT-EXPRESSED KINETOCHORE PROTEIN,
; TITLE OF INVENTION: AND METHODS OF USE
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: DANN, DORFMAN, HERRELL AND SKILLMAN
; STREET: 1601 MARKET STREET, SUITE 720
; CITY: PHILADELPHIA
; STATE: PA
; COUNTRY: USA
; ZIP: 19103-2307
; COMPUTER READABLE FORM: disk
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/353,700
; FILING DATE: 09-DEC-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: REED, JANET E.
; REGISTRATION NUMBER: 36,252
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (215) 563-4100
; TELEFAX: (215) 563-4044
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3248 amino acids

TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: HUMAN
US-08-353-700-1

Query Match 2.3%, Score 228; DB 1; Length 3248;
Best Local Similarity 18.1%; Pred. No. 4.1e-06;
Matches 376; Conservative 346; Mismatches 774; Indels 578; Gaps 93;

QY 61 NPTSDYQSVKALLNGKTFDPKSSSEFTDFVSKPDEL--TNGRTVLETPKKYQVVISFSP 119
DB 1066 NRKNELEQLKEAF-----AKEHQEFUTKLAFAEERQNLMELETVOQALRSEWTD 1116
QY 120 EDDKERFRLGPHLKEKLEDGNIQAQSAATFYLLPLDMPKAAALGQSYIVDKNFNNLI--- 176
DB 1117 NQN-----NSKSEAGGLKQE---IMTLKEQNK-----MQKEVNDLLQEN 1153
QY 177 -----IHPISNFAQSIKPLALTRSSDFTAKLQ---FNNODELVW-----YLEK 218
DB 1154 EQLMKVMKTKHECONLESEPIRNSVKERESE---RNQC�FKPQMDLEVKELSLDSYNAQ 1209
QY 219 FFDLEALKAN-----IRLOTADFSFEKGNLVD-----PFV 248
DB 1210 LVQLEAMLRNKLKLOSEKEKECLOHELOTIRGDLTSLNLOMQOSEISGLKCEIDAE 1269
QY 249 YSFTIRNPQNKQEWASDLNQOKTVRLYLRTFFS-----POAKTILKDYKYKDETFLSSID 303
DB 1270 EKYISGPHE---LSTQSDNAHLQCSLQTTMKNLEKEICEILOAEKYELVTELNDNR 1325
QY 304 LKASNGTSLFANE-----NDLKQDLVDLLDVSDYFQGQ--SETTITSNQVKPVP 351
DB 1326 SECITATRKMAEEVGGKLLNEVKILNDDSGLIHGL--VEDIPGEGFEGQNEQHPVSLAP 1383
QY 352 ASERSLKDRVKFKDQKQPRI-----EKFSLYEYDALSPYQLOELVSKPNSIKDLVNATL 407
DB 1384 LDESNSYEHLTLSKDEQVMPHAELOEKPLSIQSSEKILHDOHQCOMSSQMSLOQYVDSLK 1443
QY 408 ARNLRFSLGKYNF---LFDOLASHLDYFL--VSKAKIKQSSITKCL-----FIELP 454
DB 1444 AENLVLTSLNRFQGLVKEMQLGLEGLVPSLSSCVPDSSSLSSGDSFYRALLBQT 1503
QY 455 IKISLKSILGDBPN---IKTLFEKVTFKLDFNPRDVEIEKAFGL-----LYPG 501
DB 1504 GDMSSLSNLEGAVSANQCSVDVFCSSLQTYVDSLKAENLVLTSLNLRNFQGLVKEMQLG 1563
QY 502 VNEELEQAKAQRASFEKEKSKGLKEFSQ---QKEENSKAINNOEGLE-----547
DB 1564 LEEGLVPSLSSCVPDSSSLSSGDSFYRALLBQTGDMSSLSNLEGVVSANQCSVDVDF 1623
QY 548 ---EDDNITERLPENSPIOYQENAGLGAEPDPYMIKDVQNRYYLAKSQIOBELIKAK 603
DB 1624 CSSLOEENLTRETSPAPAKVBEHLSL---CEVYR---QSLKLEKESQGIKMK 1675
QY 604 DYTAKLILSNRHTYNIISRLKEQLFDVNPRIPSRDIEKAKFVLDTKKNKYWQIYSSA 663
DB 1676 BIQLEQLLSERQELDCLR-KOYL-----SENEQWQOKLTSV 1712
QY 664 SPVQNKWSLFGYYRYLLGLDPKOTIHELVLKQAGLOQFEGYENLPDFNLEDLKNRI 723
DB 1713 TLEMESSK-----LAAEKKQTEQLSLEL-EVARLQLOGL-----DLSRSLLGIDT 1756
QY 724 KTLPLFSQKDNFKLSLLDFNNYDGEI--KAEFGFLPLFLPKELRN-----767
DB 1757 EDAIQGNES-----CDISKEHTSETTERTPKHDVHQICDKDAQDNLNDIEKITGTAV 1811
QY 768 --SSNSGGSQNSNPWE-----QETISQ--FKDON-----LSNQDLAQFSTK 806
DB 1812 KPTGCSGEGQPDNTNYPPEGDKTQGSSECSISLSEFSGFNALVPMDFLGNQEDIHNLQR 1871

QY 1438 ERSIGVPMWSTGLD-----GYIGSEQTKDGTSSSQKGF-- 1472
 Db 1629 ---LAIWSEMAVNGLSLIEIALAKGLKRNPNLEYISKTSTRIGLNLISIGFDIY 1685
 QY 1473 -----QDFI-----OALG-----LKNTYHKGKLG- 1491
 Db 1686 NIYDNFSRISVEKNEKRRIYIVNGSLAIVGSLVTLGVSIAMLAGSTVAGPIGIVAGAVI 1745
 QY 1492 -LSIRIFDPCNELAKIKDASNKG--EKKLSKYDLF-----KN---YLNEYEKK 1535
 Db 1746 ALAISY--NAARLIEEAKKIGFTPLBELNNGFYAFLMGDILPCKKNEIVLETISQL 1802
 QY 1536 SPKIAKGTNIHPDQKE-----YPNPNQKLPENYLNVLNQPMKVTLYNSDDFI 1584
 Db 1803 ETWIAQ--NAQEHLEIKKEQEPHSYFYSNKQLYEEHHYKVIPIRLKCTL-----DNI 1855
 QY 1585 TNLFVPEGSDRSGTKLKQVIOKQNNNVADWGA---YLT----- 1623
 Db 1856 LNPF-----GEIVVQVAKNLSKEEAKIAALSHYLSVTERTEYKYVPKBAIATD 1905
 QY 1624 --FWYDKNIITNQPNVITANIADVFIKVKLESDNTKL-----IAPNITQWPNISG-- 1673
 Db 1906 EILIFDMDFYVQNLNRYTINFA--FEDDNPVFEVDVDSIFDKIRTTKEKTWQLSSGEL 1963
 QY 1674 -----SKEKFYKPT-----VFFGNWENENSSMNSQAQPTTWEKIREGFALQALKSSF 1720
 Db 1964 TDDLKSTAEKQYFEPFSWRENELFYFTYNGDDIIAAPLITQNIQFVHNG-----TKRLSG 2019
 QY 1721 DQKTRTF-VLTNNAP 1734
 Db 2020 GEXDDTENVTSESP 2034

Search completed: October 28, 2005, 21:09:11
 Job time : 132 secs

QY 633 PRIPSSRDIEAKFVLDKTERKNKYQWYSSASPVFNKWSLFGYYRYLLGLDPKQTIHEL 692
 Db 701 -----KIVSLIDTLRGNY----- 714
 QY 693 VKLGQKAGLOEGEVNLPDNLNIRIKTLPFSQKONFKLSLLDENNYDEIKAP 752
 Db 715 -----FDKITASFSIID-----EILFNQ--NNENISLVDILNNOVKQKLEDK 754
 QY 753 BFGLPLFLPKE-----LRNNSNGSGSQNSPWEQEIISQF-----K 790
 Db 755 LFYSDLLYGEKELALAKNSLVAALFYRMVAVNEINEGYTGISNFIVEQIILNPFILDK 814
 QY 791 DONLS--NODQLAOFST--KIWEKIIG--DNE-----FQNNRLQYKLLKDIQESWIN 838
 Db 815 NRNVGVGYDVTINFKNKYKALKEIIDIGIDISEIKNIFLLDDENRHLKVTYQRLSEYN 874
 QY 839 KTRDNLYWYLGDKLVKPKNLEAKFQISN--LOELLTAFVTSAAALSNNMNYQDSG 895
 Db 875 NKYVKNLHLIEDEID-----SIKKIDNTGIYRDLYLEFFSNHOFNR----- 917
 QY 896 AKSTIIFEEIAEL-----DPKV-----KEKVGADVVQL 923
 Db 918 ----VIHEIAQLESYFNNNYQERNHGYHPSNFYINENDPLTLEKSPNQKNNEIDGSL 973
 QY 924 KPH-----YAGFDN-----AGKFNQEVIRSSRTIYLTSGSKSLEA 962
 Db 974 LFNESKENFYFLVDNPFITRDFVDNIIIESISGVYQDDIV-----IYFYNGVKS----- 1023
 QY 963 DTIDOLNQAQKAP--LGLOSFYLD-----TERFGVFOKLATSLAVOHKOKE 1007
 Db 1024 ---EQLTYLKDQKEIGR---FLDYCLKHEIRVIAAGNEEYFP-----ATNALVKQKSRV 1072
 QY 1008 KTLPKLANNDGYTLIHDKLKKBVPIQIISSEPKDMFEGKLANQSGQNNVNSTP-----G 1062
 Db 1073 ESLQVNI-----LQHQFVGERTIVFANKETLFSYQSGHLFIEGIAQLNMPITYVTDNS 1126
 QY 1063 STIESPYF-----STNFOEDADLDQDG--- 1084
 Db 1127 VILENEYIMVKPTTESQFIGHPELLADTGASNTLLTETVPNEPTVNSRQRVDIENKNYAA 1186
 QY 1085 ---QDSSRGQNSLNOEAGL--LKOKLAILGN-----QFIOYQONDKEISFEIIN 1132
 Db 1187 ETYQFVSHNYTRYRDNQEPKGYQNEISTIPPDYSDNMTIADILNYIQLNRYQINYAOLG 1246
 QY 1133 --VEKVSLSFRVEFKLAKTLEDNGKTIRVLSDEMTSLIVNTTIBK--TPEMSAAVEPVEDT 1189
 Db 1247 AIIRKVDQIN--LKXHKELAEFSDKII-----NHDLIDNAIEKYLVELSAIENTSDN 1298
 QY 1190 KWVEQ-----YDPRTPLAATKTVLKFQDQIPVDGSG-----NISDKWLASIPL----- 1233
 Db 1299 SLIKRNLISQLYDPS-----VNKEFNQFLMDGTNPEOMQOEITNKDYGNYSLTERTY 1349
 QY 1234 ----VHQOMLRLSPVWKTIRELG-----LKTEQOQOQOQOQOQOQO 1270
 Db 1350 QVIEIHH-AIYDTPHLIDLSLSKVRNLNAPFDEGNKGLSVLLSKISSFELYKETINN 1408
 QY 1271 POKKAV--AKEELETYNPKDEFNINLPL-----TKAHLRLSLNLVNN----- 1311
 Db 1409 LKKLAVISQDEHSALSPEALNKLNEFKGEVNLATSEDSSTISSKVEINNVLIDKRL 1468
 QY 1312 -----DPNYKIED--LKVIKNEAGHQLA-----FS 1335
 Db 1469 CYLGAKINGESSIDAIDLNOIENWDKKITPDQOH--LNDYFLSVTGNKEKKEVIALRLYL 1527
 QY 1336 LRANNIKRLMNTPIITADY-----NPFYFYNEDWRSIDKYLNNKGNVSSHQ 1382
 Db 1528 NKQDKIKYFLPSHWSRADYVAHAERLWKIALGSKOYNQODWENL--RIFSLK--IPRHR 1584
 QY 1383 ---QAAGGNGGSLIQRLKN--IKPETETPALIALKDRNNTNLSNYSDKIIMIKPKYLV 1437
 Db 1585 VISKIGYANISFGIWSQINSFTMLAEQLNPLTTPQEHRETINY----- 1628

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 28, 2005, 20:53:38 ; Search time 243 Seconds

(without alignments)
3959.658 Million cell updates/sec

Title: US-10-607-631-8

Perfect score: 9732
Sequence: 1 MNKKSTLLATAAIIIGST.....TNNAFNVFKEPNISKIVE 1879

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1612378 seqs, 512079187 residues

Total number of hits satisfying chosen parameters: 1612378

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Uniprot 03.*

1: uniprot_sprot.*

2: uniprot_trembl.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	9701	99.7	1879	Q71A34	Q71A34 mycoplasma
2	9435.5	97.0	1878	Q71A35	Q71A35 mycoplasma
3	9402	96.6	1871	Q719M1	Q719M1 mycoplasma
4	2816	28.9	560	Q9KGX9	Q9KGX9 mycoplasma
5	2770	28.5	560	Q9KGX7	Q9KGX7 mycoplasma
6	523	5.4	1405	Q8KWK1	Q8KWK1 mycoplasma
7	502.5	5.2	1427	Q8G122	Q8G122 mycoplasma
8	473.5	4.9	1296	Q9KGX3	Q9KGX3 mycoplasma
9	410.5	4.2	1078	Q6R5H9	Q6R5H9 mycoplasma
10	407	4.2	1037	Q6R518	Q6R518 mycoplasma
11	405	4.2	5767	Q8I525	Q8I525 plasmodium
12	404	4.2	1063	Q6R510	Q6R510 mycoplasma
13	399	4.1	1077	Q6R514	Q6R514 mycoplasma
14	395.5	4.1	2965	Q6R550	Q6R550 plasmodium
15	394	4.0	1099	Q49560	Q49560 mycoplasma
16	393	4.0	1092	Q07132	Q07132 mycoplasma
17	391.5	4.0	1092	Q6R5H7	Q6R5H7 mycoplasma
18	391.5	4.0	1108	Q49542	Q49542 mycoplasma
19	390.5	4.0	1098	Q6R516	Q6R516 mycoplasma
20	390	4.0	2740	Q7RFB2	Q7RFB2 plasmodium
21	390	4.0	2976	Q7RFB2	Q7RFB2 plasmodium
22	389.5	4.0	2664	Q7RE10	Q7RE10 plasmodium
23	389	4.0	1093	Q6R5H8	Q6R5H8 mycoplasma
24	387.5	4.0	2770	Q7YUE9	Q7YUE9 plasmodium
25	387.5	4.0	2969	Q7KFP73	Q7KFP73 plasmodium
26	385.5	4.0	2261	Q7RRJ3	Q7RRJ3 plasmodium
27	384	3.9	2757	Q7RRJ9	Q7RRJ9 plasmodium
28	382.5	3.9	1098	Q6R512	Q6R512 mycoplasma
29	380.5	3.9	2977	Q6WP99	Q6WP99 plasmodium
30	380	3.9	2747	Q9BJX9	Q9BJX9 plasmodium
31	378.5	3.9	3427	Q6YA79	Q6YA79 plasmodium

ALIGNMENTS

RESULT 1

Q71A34 Q71A34 PRELIMINARY; PRT; 1879 AA.
AC Q71A34;
DT 05-JUL-2004 (Tremblrel. 27, Created)
DT 05-JUL-2004 (Tremblrel. 27, Last sequence update)
DT 05-JUL-2004 (Tremblrel. 27, Last annotation update)
DE F216 surface protein.
OS Mycoplasma hyopneumoniae.
OC Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.
OX NCBI_TaxID=2099;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=232;
RA Wilton J., Cordwell S., Minion C., Hahn M., Walker M., Djordjevic S.;
RL Submitted (AUG-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF540381; RAQ11196.1; --
SQ SEQUENCE 1879 AA; 215648 MW; 6F255658D97BB8E2 CRC64;

Query Match 99.7%; Score 9701; DB 2; Length 1879;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 1875; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY	1	MNKKSTLLATAAIIIGSTVFCTVGLASVKYRGVNPTQGVISQLGIDSVAKPSTA	60
Db	1	MNKKSTLLATAAIIIGSTVFCTVGLASVKYRGVNPTQGVISQLGIDSVAKPSTA	60
QY	61	NFTSDYQSVKALLNGKTPDKSSBFTDPVSKFDFLTNNGRVTVEIPKKYQVVISFSP	120
Db	61	NFTSDYQSVKALLNGKTPDKSSBFTDPVSKFDFLTNNGRVTVEIPKKYQVVISFSP	120
QY	121	DKERFRLGFLHKEKLEGNIAQSATKFTYLPDMPKAALGOYSYVDKNFNLLIHL	180
Db	121	DKERFRLGFLHKEKLEGNIAQSATKFTYLPDMPKAALGOYSYVDKNFNLLIHL	180
QY	181	SNFSAQSIKPLATRSSDFIAKLNQNNQDLVLYLEKFFDLKALKANRLQTADFSEK	240
Db	181	SNFSAQSIKPLATRSSDFIAKLNQNNQDLVLYLEKFFDLKALKANRLQTADFSEK	240
QY	241	GNLVDPFVYSFIRNPQNKWASDLNQDKTVRLYLRTFSPQAKTILKDYKYKDTFLS	300
Db	241	GNLVDPFVYSFIRNPQNKWASDLNQDKTVRLYLRTFSPQAKTILKDYKYKDTFLS	300
QY	301	SIDLKASNGTSLFANENDLKQDLVDLLVDVDFGQSETITSNSQVKKVPASERSLKDR	360
Db	301	SIDLKASNGTSLFANENDLKQDLVDLLVDVDFGQSETITSNSQVKKVPASERSLKDR	360
QY	361	VFKKQKQPRIEKFSLEYDALSFSYQQLVSKENSIDLVNATLARNRFSLGKYNF	420
Db	361	VFKKQKQPRIEKFSLEYDALSFSYQQLVSKENSIDLVNATLARNRFSLGKYNF	420
QY	421	LPDDLASHLDYFVLVSKAKIKQSSITTKUFIPIKISLKSILGQDQEPNITLFEKEVT	480
Db	421	LPDDLASHLDYFVLVSKAKIKQSSITTKUFIPIKISLKSILGQDQEPNITLFEKEVT	480

32 377 3.9 2269 2 Q26223 Q26223 plasmodium
33 376.5 3.9 2661 2 Q7RMS4 Q7RMS4 plasmodium
34 376.5 3.9 2749 2 Q7REV3 Q7REV3 plasmodium
35 376.5 3.9 2957 2 Q8WR56 Q8WR56 plasmodium
36 373 3.8 2849 2 Q8IHY4 Q8IHY4 plasmodium
37 372.5 3.8 2752 2 Q9BJY0 Q9BJY0 plasmodium
38 372 3.8 2070 2 Q7RBM8 Q7RBM8 plasmodium
39 371 3.8 2771 2 Q26216 Q26216 plasmodium
40 370.5 3.8 1939 2 Q25662 Q25662 plasmodium
41 369.5 3.8 2723 2 Q7RQB6 Q7RQB6 plasmodium
42 367.5 3.8 3317 2 Q8WEP8 Q8WEP8 mycoplasma
43 366.5 3.8 2093 2 Q7RJA2 Q7RJA2 plasmodium
44 364 3.7 2719 2 Q7RPU0 Q7RPU0 plasmodium
45 359.5 3.7 2699 2 Q7RLQ0 Q7RLQ0 plasmodium

```
Qy 481 PKLDNFRDVEIEKAGLLYPGVNELEOARAKAASFEKESKGLKEFSQOKEENS KAI 540
Db 481 PKLDNFRDVEIEKAGLLYPGVNELEOARAKAASFEKESKGLKEFSQOKEENS KAI 540
Qy 541 NNQGLEBDDNITERLPENSPIOYQOENAGLGASDPKPYMIKDVQNRQYLLAKSQIOBELI 600
Db 541 NNQGLEBDDNITERLPENSPIOYQOENAGLGASDPKPYMIKDVQNRQYLLAKSQIOBELI 600
Qy 601 KAKDYTKLAKLLSNRHTYNI SLRLKEQLFDVNPRI PSSRDIEKAKFVLDDTKCKNYQIY 660
Db 601 KAKDYTKLAKLLSNRHTYNI SLRLKEQLFDVNPRI PSSRDIEKAKFVLDDTKCKNYQIY 660
Qy 661 SSASVFQKNKSLFGYRYLLGLDPKQITHELVKLGQKAGLQFEGYENLPDSFNLDELKN 720
Db 661 SSASVFQKNKSLFGYRYLLGLDPKQITHELVKLGQKAGLQFEGYENLPDSFNLDELKN 720
Qy 721 TRIKTPLSQKDNFKLSLLDPFNYYDGEI KAPEFGLPLPLPKELRRNSSGGSQNSNP 780
Db 721 TRIKTPLSQKDNFKLSLLDPFNYYDGEI KAPEFGLPLPLPKELRRNSSGGSQNSNP 780
Qy 781 WEQEIISQFKQNLQNLQDLAQFS TKIWEKIIIGDENEFQNNRLOYLKLDLQESWINKT 840
Db 781 WEQEIISQFKQNLQNLQDLAQFS TKIWEKIIIGDENEFQNNRLOYLKLDLQESWINKT 840
Qy 841 RDNLTYTYLGDKLKVKPKNNLEAKPRQISNLQELLTAFTYTSAAALSNWNYYQDSGAKSTI 900
Db 841 RDNLTYTYLGDKLKVKPKNNLEAKPRQISNLQELLTAFTYTSAAALSNWNYYQDSGAKSTI 900
Qy 901 IFEETAEIDLPKVEKVGADVTOLKPHYAIGFDDNAGKQNEVIRSSSTIYLYKTSGSKL 960
Db 901 IFEETAEIDLPKVEKVGADVTOLKPHYAIGFDDNAGKQNEVIRSSSTIYLYKTSGSKL 960
Qy 961 EADTTIDQLNQAVKNAPLGLQSYLDTERFGVFOKLATSLAVQHKQKXETLPKLLNNQGYT 1020
Db 961 EADTTIDQLNQAVKNAPLGLQSYLDTERFGVFOKLATSLAVQHKQKXETLPKLLNNQGYT 1020
Qy 1021 LIHDKLKKPVPQIPISSPEKQWFEGLKMQNGOSQNVNVTFTGSIIESPYFTNFQEDADL 1080
Db 1021 LIHDKLKKPVPQIPISSPEKQWFEGLKMQNGOSQNVNVTFTGSIIESPYFTNFQEDADL 1080
Qy 1081 DQGGDDSRQNSLNDQESAGLLKQKLAILLGNQFIQYQONDKEIEPIEIIINVEKVSLS 1140
Db 1081 DQGGDDSRQNSLNDQESAGLLKQKLAILLGNQFIQYQONDKEIEPIEIIINVEKVSLS 1140
Qy 1141 FRVEPKLAKTLEDNGKTIIRVLSDETMSLIVNTTIEKTPMSAVPEVFTKWEQYDPRTP 1200
Db 1141 FRVEPKLAKTLEDNGKTIIRVLSDETMSLIVNTTIEKTPMSAVPEVFTKWEQYDPRTP 1200
Qy 1201 LAAKTKFVLKFKDQIPVDGSGNISDKWLASIPLVIIHQQMLRLSPVVKTIIRGLKTEQQQ 1260
Db 1201 LAAKTKFVLKFKDQIPVDGSGNISDKWLASIPLVIIHQQMLRLSPVVKTIIRGLKTEQQQ 1260
Qy 1261 QOQOQOQOQOQKAVRKEELETYNPKDEFNIIINPLTKAHLRLTSLNVNNDPNYKIEDL 1320
Db 1261 QOQOQOQOQOQKAVRKEELETYNPKDEFNIIINPLTKAHLRLTSLNVNNDPNYKIEDL 1320
Qy 1321 KVIKNEAGDHQALFSLRANNIKRLMNTPIITADYNPFYVNEDEWRSIDKYLNNKGNVSSH 1380
Db 1321 KVIKNEAGDHQALFSLRANNIKRLMNTPIITADYNPFYVNEDEWRSIDKYLNNKGNVSSH 1380
Qy 1381 QOQAAAGGQSGSLIQRLLNKNIKPETFTPALIALKDRNNTNLSNYSDKIIMIKPKYLVERS 1440
Db 1381 QOQAAAGGQSGSLIQRLLNKNIKPETFTPALIALKDRNNTNLSNYSDKIIMIKPKYLVERS 1440
Qy 1441 IGVPHSTGLDGYIGSEQTKDGTSSSSQKGFQDFIQALGLKNTEYHGKGLGSLRIFDPG 1500
Db 1441 IGVPHSTGLDGYIGSEQTKDGTSSSSQKGFQDFIQALGLKNTEYHGKGLGSLRIFDPG 1500
Qy 1501 NELAKIKDASNNKGEKLLKSYDLFPKNYLNVEYKSKPKIAKGWNIHDPQKEYPNQKL 1560
Db 1501 NELAKIKDASNNKGEKLLKSYDLFPKNYLNVEYKSKPKIAKGWNIHDPQKEYPNQKL 1560
```

```
Qy 1561 PENYLNILVNLQPKWKTLYNSSDFTITNLFVPEPESDRSGTGKTKQVIOQVNNNNYADWGA 1620
Db 1561 PENYLNILVNLQPKWKTLYNSSDFTITNLFVPEPESDRSGTGKTKQVIOQVNNNNYADWGA 1620
Qy 1621 YLTFWYDKNIITNQPNVITANIADVFIKOVKELEDNDTKLAPNITQWPNISGSKFPYK 1680
Db 1621 YLTFWYDKNIITNQPNVITANIADVFIKOVKELEDNDTKLAPNITQWPNISGSKFPYK 1680
Qy 1681 PTVPFQGNWENESSMSQAOTPTWEKIREGFALQALKSSFDQKTRTFVLTNAPLPLWKY 1740
Db 1681 PTVPFQGNWENESSMSQAOTPTWEKIREGFALQALKSSFDQKTRTFVLTNAPLPLWKY 1740
Qy 1741 GPLGFQGNPFKTDQWRLVFQNDNQIAALRVQEQDRPEKSSDKQKWKFKFVPIPEE 1800
Db 1741 GPLGFQGNPFKTDQWRLVFQNDNQIAALRVQEQDRPEKSSDKQKWKFKFVPIPEE 1800
Qy 1801 MPNSGNIRFVGVMQIOQPNITLWLPVINSVITYDFYRGTDGSDNDVANLNVAPQVKTIAFT 1860
Db 1801 MPNSGNIRFVGVMQIOQPNITLWLPVINSVITYDFYRGTDGSDNDVANLNVAPQVKTIAFT 1860
Qy 1861 NNAFNNVFKEFNISKKIVE 1879
Db 1861 NNAFNNVFKEFNISKKIVE 1879

RESULT 2
Q71A35 PRELIMINARY; PRT; 1878 AA.
AC Q71A35, 2004 (TrEMBLrel. 27, Created)
DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE P216 surface protein.
OS Mycoplasma hyopneumoniae.
OC Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.
OX NCBI_taxid=2059;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=J;
RA Wilton J., Cordwell S., Minion C., Hahn M., Walker M., Djordjevic S.;
RL Submitted (AUG-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF540380; AAQ11195.1; -.
SQ SEQUENCE 1878 AA; 215617 MW; B6FDA7B8B2DB9037 CRC64;

Query Match 97.0%; Score 9435.5; DB 2; Length 1878;
Best Local Similarity 97.0%; Pred. No. 0;
Matches 1827; Conservative 17; Mismatches 29; Indels 11; Gaps 3;

Qy 1 MNKKSTLLATATAAIIIGSTVGTGVGLASKVKYRGVNPQTGGVLSQLGLSDSVAFKPSIA 60
Db 1 MNKKSTLLATATAAIIIGSTVGTGVGLASKVKYRGVNPQTGGVLSQLGLSDSVAFKPSIA 60
Qy 61 NFTSDYQSVKALLNGKTFDPKSEFTDPVSKPFDLTNNGRTVLEIPKKYQVVISSEFSPE 120
Db 61 NFTSDYQSVKALLNGKTFDPKSEFTDPVSKPFDLTNNGRTVLEIPKKYQVVISSEFSPE 120
Qy 121 DDKERFRLGPHLKEKLEDGNIQAQATFIYLLPLDMPKAALGOYSYIVDKNFNNLIHPL 180
Db 121 DDKERFRLGPHLKEKLEDGNIQAQATFIYLLPLDMPKAALGOYSYIVDKNFNNLIHPL 180
Qy 181 SNFSAQSIKPLALTRSSDPIAKLNQFNQDELWVYLEKPFDEALKANIRLOTADFSFEK 240
Db 181 SNFSAQSIKPLALTRSSDPIAKLNQFNQDELWVYLEKPFDEALKANIRLOTADFSFEK 240
Qy 241 GNLVDPFVYSFIRNPQKEWASDLNQDKTQVRLYLRTFSPQAKTILKDYKYKDETFLS 300
Db 241 GNLVDPFVYSFIRNPQKEWASDLNQDKTQVRLYLRTFSPQAKTILKDYKYKDETFLS 300
Qy 301 SIDLKASNGTSLFANENDLKQDLVDLLDVSDFGQSGSETITSNSQKVPASRSRKDR 360
Db 301 SIDLKASNGTSLFANENDLKQDLVDLLDVSDFGQSGSETITSNSQKVPASRSRKDR 360
Qy 361 VKFKDQOKPRIEKFSIYEYDALSFYSQQLVSKPNSIKDLVNLATLARNLRFSLGKNYF 420
```



```
Db 361 VFPKDDQKPRTEKSLYEDALSFYSQELVSKPNSIKDJVNATLAKNLFSLGKNF 420
QY 421 LPDDLASHLDYFLVSKAKIKOSSITKFLFELPKISLKSSILGDQEPNIKTLEKEVT 480
Db 421 LPDDLASHLDYFLVSKAKIKOSSITKFLFELPKISLKSSILGDQEPNIKTLEKEVT 480
QY 481 FKLDNFRDVEIEKAFGLIPGVNHELEQARKORASFEKEKSKGLKEFSQOKEENSKAI 540
Db 481 FKLDNFRDVEIEKAFGLIPGVNHELEQARKORASFEKEKSKGLKEFSQOKEENSKAI 540
QY 541 NNOEGLEEEDNITERLPENSPLOYQOENAGLAGSPDKPMIKOVQORYLAKSQIQELI 600
Db 541 NNOEGLEEEDNITERLPENSPLOYQOENAGLAGSPDKPMIKOVQORYLAKSQIQELI 600
QY 601 KAKDYTKLAKLSNRHTYINISRLKEQLFDVNPRIPISSRDIIEKAFVLDKTEKNKYWQIY 660
Db 601 KAKDYTKLAKLSNRHTYINISRLKEQLFEVNPRIPISSRDIENAKFVLDTKEKNKYWQIY 660
QY 661 SSASPVPQNKWSLFGYRYRLLGLDPKQTHLVLKQKAGLOPEGYENLPSPFNLEDLKN 720
Db 661 SSASPVPQNKWSLFGYRYRLLGLDPKQTHLVLKQKAGLOPEGYENLPSPFNLEDLKN 720
QY 721 IRIKTPLESQKPKLSLLDFNNYDGEIKAPFGLPLPKELRNRSSNGSSQNSNP 780
Db 721 IRIKTPLESQKPKLSLLDFNNYDGEIKAPFGLPLPKELRNRSSNGSSQNSNP 780
QY 781 WEQEIISQFQDNLSQDQLAQFSTIKWEKIIIGDENEFDQNNRLOKLLKDLQESWINT 840
Db 781 WEQEIISQFQDNLSQDQLAQFSTIKWEKIIIGDENEFDQNNRLOKLLKDLQESWINT 840
QY 841 RDNLYTYIGDKLVKPKNNLEAKFROI SNLQELLTAFYTSAAALSNNNYQDSGAKSTI 900
Db 841 RDNLYTYIGDKLVKPKNNLEAKFROI SNLQELLTAFYTSAAALSNNNYQDSGAKSTI 900
QY 901 IFBEIAELDPKVEKGVADYVQLKEHVAIGFDNAGKFNQEVIRSSRTIYLTSGSKSL 960
Db 901 IFBEIAELDPKVEKGVADYVQLKEHVAIGFDNAGKFNQEVIRSSRTIYLTSGSKSL 960
QY 961 EADTIDQLNQAVNAPLGLQSLDTERFGVFKLATSIAVQHKQKELPKKLNDGYT 1020
Db 961 EADTIDQLNQAVNAPLGLQSLDTERFGVFKLATSIAVQHKQKELPKKLNDGYT 1020
QY 1021 LIHDKLKKPVIPOISSPEKDWTEGKLNQNGQSNVSTFGSIIIESPYSTNFQEDADL 1080
Db 1021 LIHDKLKKPVIPOISSPEKDWTEGKLNQNGQSNVSTFGSIIIESPYSTNFQEDADL 1080
QY 1081 DQGGDDSRQGNLSNDQEAAGLKQKLAIIILGNQFIQYQOONDKETEFELINVEKYSLS 1140
Db 1081 DQGGDDSRQGNLSNDQEAAGLKQKLAIIILGNQFIQYQOONDKETEFELINVEKYSLS 1140
QY 1141 FRVEFKLAKTLEONGKTIIRVLSDETMSLI VNTTIEKTPEMSAPVFPDTKWEQYDPRTP 1200
Db 1141 FRVEFKLAKTLEONGKTIIRVLSDETMSLI VNTTIEKAPESAAPEVFTDKWEQYDPRTP 1200
QY 1201 LAATKTVLKFQKQIPVDGSGNISDKWLASIPLVIHQOMLRLSPVVKTIIRLGLKTEQQQ 1260
Db 1201 LAATKTVLKFQKQIPVDASGNISDKWLASIPLVIHQOMLRLSPVVKTIIRLGLKTEQQQ 1260
QY 1261 QQQQQQQQQQPOKAVRKEBELETYNPKDEFNLIPLTKAHRLTLSLVNNDPNYKIEDL 1320
Db 1261 QQQQQQ-----QKAVRKEBELETYNPKDEFNLIPLTKAHRLTLSLVNNDPNYKIEDL 1315
QY 1321 KVIKNEAGDHQAFSLRANNIKRLMNTPIITPADYNPFYFNEWRSDIKYLNKGNVSS- 1379
Db 1316 KVIKNEAGDHQAFSLRANNIKRLMNTPIITPADYNPFYFNEWRSDIKYLNKGNVSSQ 1375
QY 1380 ----HQQAAGGNGSGSLIQRLNKQIIPETFPALIALKDRNNTLSNYSDKIIIMPKY 1435
Db 1376 QQQQQQQQPPGGNGSGSLIQRLNKQIIPETFPALIALKDRNNTLSNYSDKIIIMPKY 1435
QY 1436 LVERSIGVPMSTGLDGYIGSGTKDGTSSSQQKGFQDQFTQALGLKNTYHGLGLSIR 1495
```

```
Db 1436 LVERSIGVPMSTGLDGYIGSGTKDGTSSSQ--QKRFKODFTQALGLKNTYHGLGLSIR 1494
QY 1496 IFDPGNEELAKIKDASNKKGEKILKSYDLFRNYLNEVEKSPKIAKGWNTNHPDOKEYPN 1555
Db 1495 IFDPGNEELAKIKDASNKKGEKILKSYDLFRNYLNEVEKSPKIAKGWNTNHPDOKEYPN 1554
QY 1556 PNQKLPENYLNVLNQPWKVTLVNSSDFTITMLFVEPEGSGRSGTGKLVQVIOKQVNNNYA 1615
Db 1555 PNQKLPENYLNVLNQPWKVTLVNSSDFTITMLFVEPEGSGRSGTGKLVQVIOKQVNNNYA 1614
QY 1616 DWGSAYLTFWYDKMIITNQPNVITANIADVPIDKVKLELWNTKLIAPNITOWWPNISGSK 1675
Db 1615 DWGSAYLTFWYDKMIITNQPNVITANIADVPIDKVKLELWNTKLIAPNITOWWPNISGSK 1674
QY 1676 EKFKYPTVPFGNWNENSSNMQOQTPTEKIREGFALQALKSSFDQKTRTFLVLTNAPL 1735
Db 1675 EKFKYPTVPFGNWNENSSNMQOQTPTEKIREGFALQALKSSFDQKTRTFLVLTNAPL 1734
QY 1736 PLWKYGPGLGFQNGPNFKTQDWRLVFNQDNDNQIAALRVQEQDRPEKSSDKQKWKFKV 1795
Db 1735 PLWKYGPGLGFQNGPNFKTQDWRLVFNQDNDNQIAALRVQEQDRPEKSSDKQKWKFKV 1794
QY 1796 VIPEMFNSGNIRFVGVMQIOGPNLTMLPVINSSVIYDFYRGTTGDSNDVANLVAPWQVK 1855
Db 1795 VIPEMFNSGNIRFVGVMQIOGPNLTMLPVINSSVIYDFYRGTTGDSNDVANLVAPWQVK 1854
QY 1856 TIAFTNNAFNNVPKFNISKKIVE 1879
Db 1855 TIAFTNNAFNNVPKFNISKKIVE 1878

RESULT 3
Q719M1 PRELIMINARY; PRT: 1871 AA.
ID Q719M1;
AC Q719M1;
DT 05-JUL-2004 (TREMELrel. 27, Created)
DT 05-JUL-2004 (TREMELrel. 27, Last sequence update)
DT 05-JUL-2004 (TREMELrel. 27, Last annotation update)
DE P216.
OS Mycoplasma hyopneumoniae.
OC Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.
OX NCBI_TaxID=2099;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Beaufort;
RA Wilton J.L., Cordwell S.J., Minion F.C., Guerreiro N., Hahn M.,
Walker M.J., Djordjevic S.P.;
RL Submitted (AUG-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF541877; AAQ11369.1; -.
SQ SEQUENCE 1871 AA; 214796 MW; CAAAA5PPE1CCE9BB CRC64;

Query Match 96.6%; Score 9402; DB 2; Length 1871;
Best Local Similarity 96.7%; Pred. No. 0;
Matches 1819; Conservative 22; Mismatches 28; Indels 12; Gaps 3;

QY 1 MNKKSTLLATAAAIIIGSTVGVGLASKYKVRGVNPTQGVISQGLIDSVAFKPSIA 60
Db 1 MNKKSTLLATAAAIIIGSTVGVGLASKYKVRGVNPTQGVISQGLIDSVAFKPSIA 60
QY 61 NFTSDYQSVKALLNGKTFDPKSSBFTDFVGFDFLTNNGRVLEIPKKYQVVISFSPSE 120
Db 61 NFTSDYQSVKALLNGKTFDPKSSBFTDFVGFDFLTNNGRVLEIPKKYQVVISFSPSE 120
QY 121 DKERFRLGFLHKEKLEDGNIASQATKFIYLAFLQMPKALGOYSYIVDKNFNNLIHPL 180
Db 121 DKERFRLGFLHKEKLEDGNIASQATKFIYLAFLQMPKALGOYSYIVDKNFNNLIHPL 180
QY 181 SNFSAQSIKPLATIRSSDFIAKLNQFNQDELWVYLEKFEFDEALKANIRLOTADSFSEK 240
Db 181 SNFSAQSIKPLATIRSSDFIAKLNQFNQDELWVYLEKFEFDEALKANIRLOTADSFSEK 240
QY 241 GNLVDPFVYSFIRNPQNKQEWASDLNQDQKTVRLYLRTBFSPQAKTILKDYKYKDTFSL 300
```

Db 241 GNLVDPFVYFIRNPQOKEWASDLNQDQKTVRLYLRTFSPQAKTILKDYKYDDEFLS 300
Qy 301 SIDLKASNGTSLFANENDLKQDLDVLDVSDYFGQSQSETTTSNSQVKPVPASERSLKDR 360
Db 301 SIDLKASNGTSLFANEDDLKQDLDVLDVSDYFGQSQSETTTSNSQVKPVPASERSLKDR 360
Qy 361 VKFKDQDQKPRIEKSLVEYDALSPYSQLEULVSPNSIKOLVNATLARNRFLSGKYNF 420
Db 361 VKFKDQDQKPRIEKSLVEYDALSPYSQLEULVSPNSIKOLVNATLARNRFLSGKYNF 420
Qy 421 LPDDLASHLDVYFLVSKAKIKOSSITTKLFTLPIKISLKSILGDQEPNKTLPKEVVT 480
Db 421 LPDDLASHLDVYFLVSKAKIKOSSITTKLFTLPIKISLKSILGDQEPNKTLPKEVVT 480
Qy 481 FKLDNFRDVEIEKAFGLLYPGVNEELEQARKAQRASFEKESKGLKEFSSQKEENSXAI 540
Db 481 FKLDNFRDVEIEKAFGLLYPGVNEELEQARRDORASLEKAKKGLKEFSSQKQDENLXAI 540
Qy 541 NNQGLEEDDNIATERLPENSPITQYQOENAGLSPDKPYMIKDVQONQRYYLAKSQIOBELI 600
Db 541 NNQGLEEDDNIATERLPENSPITQYQOENAGLSPDKPYMIKDVQONQRYYLAKSQIOBELI 600
Qy 601 KAKDYTKLAKLSNRHTYNISLRLEQOLFQDVNPRIPSSRDIEKAFVLDKTEKNKYQOY 660
Db 601 KAKDYTKLAKLSNRHTYNISLRLEQOLFQDVNPRIPSSRDIEKAFVLDKTEKNKYQOY 660
Qy 661 SSASPVFQNKSLFGYRYLLGLDPKQTHIELVKLGQKAGLQFEGYENLPSPDFNLEDLKN 720
Db 661 SSASPVFQNKSLFGYRYLLGLDPKQTHIELVKLGQKAGLQFEGYENLPSPDFNLEDLKN 720
Qy 721 IRIKTPLSQKDNFKSLDLLDFNYYDGEIKAPEFGLPLFLPKELRRNSNSGGGQNSNSP 780
Db 721 IRIKTPLSQKDNFKSLDLLDFNYYDGEIKAPEFGLPLFLPKELRRNSNSGGGQNSNSP 780
Qy 781 WEQEIISQFQKQNLQNOQLAQFSKIWEKIIGDENEFQNNRLOYLKDLQESWINKT 840
Db 781 WEQEIISQFQKQNLQNOQLAQFSKIWEKIIGDENEFQNNRLOYLKDLQESWINKT 840
Qy 841 RDLNLYTWYLGDKLVKPKNNLEAKFRQISNLQELLTAFTYSAALSNMNNYYQDSGAKSTI 900
Db 841 RDLNLYTWYLGDKLVKPKNNLEAKFRQISNLQELLTAFTYSAALSNMNNYYQDSGAKSTI 900
Qy 901 IPEETAEIDLPKVEKVGADVYQKPHYAIGPDDNAGKQNEVIRSSRTIYIKTSGSKL 960
Db 901 IPEETAEIDLPKVEKVGADVYQKPHYAIGPDDNAGKQNEVIRSSRTIYIKTSGSKL 960
Qy 961 EADTTIDQLNQA VKNAPLGIQSPYLDTERFGVFPQKATSLAVQHKQKEXTLPKLLNNDGYT 1020
Db 961 EADAIDQLNQA VKNAPLGIQSPYLDTERFGVFPQKATSLAVQHKQKEXTLPKLLNNDGYT 1020
Qy 1021 LIHDKLKKPVIPIQISSPEKQWFEGLKQNGSQSNVNVSTFGSIIESPYFTNFQEDADL 1080
Db 1021 LIHDKLKKPVIPIQISSPEKQWFEGLKQNGSQSNVNVSTFGSIIESPYFTNFQEDADL 1080
Qy 1081 DQGGQDSSQNGNSLDNQEAGLLKQKLAILLGNQFIQYQQNDKIEIEFEIINVEKVSLS 1140
Db 1081 DQGGQDSSQNGNSLDNQEAGLLKQKLAIXLGNQFIQYQQNDKIEIEFEIINVEKVSLS 1140
Qy 1141 FRVEFKLAKTLDNGKTIIRVLSDETMSLIVNTTIEKTPEMSAPVFPDVKWVQYDPRTP 1200
Db 1141 FRVEFKLAKTLDNGKTIIRVLSDETMSLIVNTTIEKAPEMSAPVFPDVKWVQYDPRTP 1200
Qy 1201 LAAKTKFVLKFKDQIPVDGSGNISDKWLASIPLVIHQQMLRSLPVVKTIRELGLKTEBOQ 1260
Db 1201 LAAKTKFVLKFKDQIPVDASGNISDKWLASIPLVIHQQMLRSLPVVKTIRELGLKTEBOQ 1260
Qy 1261 QOQOQOQOQOQKAVRKEELETYNPKDEENILNPLTKAHLTLNLSLVNNDPNYKIEDL 1320
Db 1261 QO-----QKAVRKEELETYNPKDEENILNPLTKAHLTLNLSLVNNDPNYKIEDL 1311
Qy 1321 KYIKNEAGDHQLEFSLRANNIKRLMNTPTITFADYNPFYFVEDWRSIDKYLNNKGNVSS- 1379
Db 1312 KYIKNEAGDHQLEFSLRANNIKRLMNTPTIXFADYNPFYFVEDWRSIDKYLNNKGNVSSQ 1371

Qy 1380 -HQOQAGGQSGGLIQRNLKNIKPFTTTPALLIALKDRNNNTNLSYSDKIIMIKPKYLVE 1438
Db 1372 QOQOQAGGQSGGLIQRNLKNIKPFTTTPALLIALKDRNNNTNLSYSDKIIMIKPKYLVE 1431
Qy 1439 RSTGVPWSTGLDGYIGSEOTKGTSSSSQKQGDODFIQALGLKNTYHKGKLSIRIFD 1498
Db 1432 RSTGVPWSTGLDGYIGSEQLKGGTSSNSQV-GPKQDFIQALGLKNTYHKGKLSIRIFD 1490
Qy 1499 PGNELAKIKDASANKGEEKLLKSYDLFPKNLYNEKEKSPKIAKGWNTNIHPDQKEYPNPQ 1558
Db 1491 PGNELAKIKDASANKGEEKLLKSYDLFPKNLYNEKEKSPKIAKGWNTNIHPDQKEYPNPQ 1550
Qy 1559 KL PENYLNVLNQPKVTLTYNSSDFITNLVPEPEGSDRGSGTKLKQVIQKQVNNNYADWG 1618
Db 1551 KL PENYLNVLNQPKVTLTYNSSDFITNLVPEPEGSDRGSGTKLKQVIQKQVNNNYADWG 1610
Qy 1619 SAYLTFWYDKNIIITNQPNVTANIADVFIKDVKELEDNTKLIAPNITQWMPNINGSKEKF 1678
Db 1611 SAYLTFWYDKNIIITNQPNVTANIADVFIKDVKELEDNTKLIAPNITQWMPNINGSKEKF 1670
Qy 1679 YKPTVFFGNWENENSSMSQAQPTWEKIREGFPALQALKSFDOKTFTFVLTITNAPLPLW 1738
Db 1671 YKPTVFFGNWENENSSMSQAQPTWEKIREGFPALQALKSFDOKTFTFVLTITNAPLPLW 1730
Qy 1739 KYGPGFQNGPNTKQDRLVFNQNDNQIAALRVQEQDRPEKSESDKDKQWIKFKVVIIP 1798
Db 1731 KYGPGFQNGPNTKQDRLVFNQNDNQIAALRVQEQDRPEKSESDKDKQWIKFKVVIIP 1790
Qy 1799 EEMFNSNIIRFVGVMQIOGPNLWLPVINSVLYDFYRGTGDSNDVANLVAPQVKTIA 1858
Db 1791 EEMFNSNIIRFVGVMQIOGPNLWLPVINSVLYDFYRGTGDSNDVANLVAPQVKTIA 1850
Qy 1859 FTNNAFNNVFEFNISKIIVE 1879
Db 1851 FTNNAFNNVFEFNISKIIVE 1871
RESULT 4
Q9KX9 PRELIMINARY; PRT; 560 AA.
ID Q9KX9 AC Q9KX9
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE YX2 (Fragment).
GN Name=YX2;
OS Mycoplasma hyopneumoniae.
OC Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.
OX NCBI_TaxID=2099;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ISU 232;
RA Liao X., Papazisi L., Geary S.;
RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF279292; AAF87781.1; -.
FT NON TER 560 560
SQ SEQUENCE 560 AA; 63805 MW; C1A89B69D7C04A18 CRC64;
Query Match 28.9%; Score 2816; DB 2; Length 560;
Best Local Similarity 99.8%; Pred. No. 1.2e-107;
Matches 559; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Qy 1 MNKKSTLLATATAAIIIGSTVFGTVGLSKVYRGVNPQTQGVISQIGLIDSVAFKPSIA 60
Db 1 MNKKSTLLATATAAIIIGSTVFGTVGLSKVYRGVNPQTQGVISQIGLIDSVAFKPSIA 60
Qy 61 NFTSDYQSVKALLNGKTFDPKSEFTDFVSKDFLTNNGRTVLEIPKKYQVWSESP 120
Db 61 NFTSDYQSVKALLNGKTFDPKSEFTDFVSKDFLTNNGRTVLEIPKKYQVWSESP 120
Qy 121 DDKERFRLGPHLKEKLDGNSIAQSAKFYILLPLDMPKAAALGOYSYIVDKNFNLIHPL 180

```

Db      121  DDKERFRLGFLHKEKLEDGNIQAQATKFIYLLPLDMPKAAALGOYSYIVDNFNLIHPL 180
QY      181  SNFSAQSIKPLATRSSDFIAKLNQFNQDELWVYLEKFFDLEALKANIRLQTADFSFEK 240
Db      181  SNFSAQSIKPLATRSSDFIAKLNQFNQDELWVYLEKFFDLEALKANIRLQTADFSFEK 240
QY      241  GNLVDPFVYSFIRNPQKQEWASDLNODQKTVRLYLRTFSPQAKTILKDYKYKDETFLS 300
Db      241  GNLVDPFVYSFIRNPQKQEWASDLNODQKTVRLYLRTFSPQAKTILKDYKYKDETFLS 300
QY      301  SIDLKASNGTSLFANENDLKQDLVDLLDVSDYFGGQSETITSNSQVQVPVPSERSLKDR 360
Db      301  SIDLKASNGTSLFANENDLKQDLVDLLDVSDYFGGQSETITSNSQVQVPVPSERSLKDR 360
QY      361  VKFKDQKQKPRIEKFSLEYDALSFYSQLQELVSKPNSIKDLVNATLARNLRFSLGKYNF 420
Db      361  VKFKDQKQKPRIEKFSLEYDALSFYSQLQELVSKPNSIKDLVNATLARNLRFSLGKYNF 420
QY      421  LFDDLASHLDYFVLVSKAKIKQSSITTKGLFIELPIKISLKSSITLGDQSPNITLFEKEVT 480
Db      421  LFDDLASHLDYFVLVSKAKIKQSSITTKGLFIELPIKISLKSSITLGDQSPNITLFEKEVT 480
QY      481  FKLDNFRDVEIEKAFGLLYPGVNEELEQAKAQRASFEKEKSKGLKEFSQOKEENSKAI 540
Db      481  FKLDNFRDVEIEKAFGLLYPGVNEELEQAKAQRASFEKEKSKGLKEFSQOKEENSKAI 540
QY      541  NNOEGLEEDDNIATERLPENS 560
Db      541  NNOEGLEEDDNIATERLPENS 560

RESULT 5
Q9KX7
ID      09KX7      PRELIMINARY;      PRT;      560 AA.
AC      09KX7;
DT      01-OCT-2000 (Tremblrel. 15, Created)
DT      01-OCT-2000 (Tremblrel. 15, Last sequence update)
DT      01-OCT-2000 (Tremblrel. 15, Last annotation update)
DE      YX2 (Fragment).
GN      Names=YX2;
OS      Mycoplasma hyopneumoniae.
OC      Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.
OX      NCBI_TaxID=2099;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN=J;
RA      Liao X., Papazisi L., Gearty S.;
RL      Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
DR      EMBL; AF279293; AAF87783.1; -.
FT      NON TER      560
SQ      SEQUENCE 560 AA; 63810 MW; 88AAC16175B7AE97 CRC64;

Query Match      28.5%; Score 2770; DB 2; Length 560;
Best Local Similarity 98.0%; Pred. No. 9e-106;
Matches 549; Conservative 4; Mismatches 7; Indels 0; Gaps 0;

QY      1  MNKKSTLLATAAAIIGSTVFGTVGLSKVKYRGVNPQTQVISOGLGLDSVAFKPSIA 60
Db      1  MNKKSTLLATAAAIIGSTVFGTVGLSKVKYRGVNPQTQVISOGLGLDSVAFKPSIA 60
QY      61  NPTSDYQSVKALLNGKTDPKSEFTDFVSKDFLTNNGRVLETPKKYQVVISSESP 120
Db      61  NPTSDYQSVKALLNGKTDPKSEFTDFVSKDFLTNNGRVLETPKKYQVVISSESP 120
QY      121  DDKERFRLGFLHKEKLEDGNIQAQATKFIYLLPLDMPKAAALGOYSYIVDNFNLIHPL 180
Db      121  DDKERFRLGFLHKEKLEDGNIQAQATKFIYLLPLDMPKAAALGOYSYIVDNFNLIHPL 180
QY      181  SNFSAQSIKPLATRSSDFIAKLNQFNQDELWVYLEKFFDLEALKANIRLQTADFSFEK 240
Db      181  SNFSAQSIKPLATRSSDFIAKLNQFNQDELWVYLEKFFDLEALKANIRLQTADFSFEK 240
QY      241  GNLVDPFVYSFIRNPQKQEWASDLNODQKTVRLYLRTFSPQAKTILKDYKYKDETFLS 300

```

```

Db      241  GNLVDPFVYSFIRNPQKQEWASDLNODQKTVRLYLRTFSPQAKTILKDYKYKDETFLS 300
QY      301  SIDLKASNGTSLFANENDLKQDLVDLLDVSDYFGGQSETITSNSQVQVPVPSERSLKDR 360
Db      301  SIDLKASNGTSLFANENDLKQDLVDLLDVSDYFGGQSETITSNSQVQVPVPSERSLKDR 360
QY      361  VKFKDQKQKPRIEKFSLEYDALSFYSQLQELVSKPNSIKDLVNATLARNLRFSLGKYNF 420
Db      361  VKFKDQKQKPRIEKFSLEYDALSFYSQLQELVSKPNSIKDLVNATLARNLRFSLGKYNF 420
QY      421  LFDDLASHLDYFVLVSKAKIKQSSITTKGLFIELPIKISLKSSITLGDQSPNITLFEKEVT 480
Db      421  LFDDLASHLDYFVLVSKAKIKQSSITTKGLFIELPIKISLKSSITLGDQSPNITLFEKEVT 480
QY      481  FKLDNFRDVEIEKAFGLLYPGVNEELEQAKAQRASFEKEKSKGLKEFSQOKEENSKAI 540
Db      481  FKLDNFRDVEIEKAFGLLYPGVNEELEQAKAQRASFEKEKSKGLKEFSQOKEENSKAI 540
QY      541  NNOEGLEEDDNIATERLPENS 560
Db      541  NNOEGLEEDDNIATERLPENS 560

RESULT 6
Q8KMK1
ID      08KMK1      PRELIMINARY;      PRT;      1405 AA.
AC      08KMK1;
DT      01-OCT-2002 (Tremblrel. 22, Created)
DT      01-OCT-2002 (Tremblrel. 22, Last sequence update)
DT      01-MAR-2003 (Tremblrel. 23, Last annotation update)
DE      Lpps protein.
GN      Name=lpps;
OS      Mycoplasma conjugativae.
OC      Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.
OX      NCBI_TaxID=45361;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN=HRC/581;
RX      MEDLINE=22464852; PubMed=12576592; DOI=10.1099/mic.0.258664-0;
RA      Bellocy L., Vilei E.M., Giacometti M., Frey J.;
RT      "Characterization of lpps, an adhesin of Mycoplasma conjugativae.";
RL      Microbiology 149:185-193(2003).
DR      EMBL; AJ318939; CAC87274.1; -.
SQ      SEQUENCE 1405 AA; 152862 MW; D9016B685051F936 CRC64;

Query Match      5.4%; Score 523; DB 2; Length 1405;
Best Local Similarity 22.5%; Pred. No. 3.7e-13;
Matches 298; Conservative 184; Mismatches 488; Indels 356; Gaps 56;

QY      1  MNKKSTLLATAAAIIGSTVFGTVGLSKVKYRGVNPQTQVISOGLGLDSVAFKPSIA 60
Db      1  MNKK--LIIGLGAAGVALFAIPVGIASITTKYDG-DPLKNVEKANSISNIVFDESI 57
QY      61  NPTSDYQSVKALL-NG-KTFDPKSEFTDFVSKDFLTNNGRVLEIPKPY---QVVIS 115
Db      58  GVNTTYASLKSKLVENGHKSDVTAAGFPD-----FTTKNQPTLTIDSENNCCIEIVK 112
QY      116  EFSPEDEKFERFLGFLHKEKLEDGNIQAQATKFIYLLPLDMPKAAALGOYSYIVDNFNLI 175
Db      113  DIIADDQDSQFKVYVYVQSKLDNDVAKSNLRVKTLSFGFLPEFSLASFAHKGQQLFKSI 172
QY      176  ITHPLSNFSA-----QSIKPLATRSSDFIAKLNQFNQDELWVYLEKFFDLEALK 227
Db      173  RPRTYGEFIATLKQAQAKNSVPLELTSSNDFAKOVNKAQSSDEAQOIWNKYFDLOSVID 232
QY      228  NTRLQADSFEEKNLVDPFVYSFIRNPQKQEWASDLNODQKTVRLYLRTFSPQAKTI 287
Db      233  RIASDTNNVFDNNGVTISKYKSLIKNPNTNNEYALPLVDNKDEVVLFLQTFESDQFKKT 292
QY      288  L-----KDYKYKDETFELSSIDLKASNGTSLFANENDLKQDLVDLLDVSDYFGGQSETIT 342
Db      293  ITNFDGDKFKHIDILRLNLKDLITS-----TKLEDKISITPFSVQDFYGRKQTNP 343

```

```
QY 343 SNSQVKVPASERSLKDRVK---FKDQOKP---RIEKFSLYEYDALSFYQLOELVSKP 396
DB 344 FTIATKEQNANILNEFVKGXIAKADFNKPGTTRVRSPLISKAISAI---TQSVEDTSVP 400
QY 397 NSI-----KOLVNAI-----LARNRFSLGKYNFLFDDLASHLD 430
DB 401 QTSSAGATGAGATTTPKTTADATLNDPTLAASDISALANQVR--TAKDFDLVDPFIEIQ 458
QY 431 YVFLVSKAKIKOSSITKK---LFTLPKISLSSILGDOEPNIKTLFEKEVTFKLDNF 486
DB 459 F-----INGSLFNKEDRDQFIFVVOGLNQDLA--KYIRFETGLENPLGSDVNF 508
QY 487 RDVEI-----EKAFGLLYFC-----VNEBLEQARKAQRASFKEKS-----KKGKKEFSQ 531
DB 509 LNFDTLDSQLDLVLYDGRPVKVGTVIVAKSNAGTIIFTKEFSTILGAFKNSLFTISK 568
QY 532 QKEENSKAINNOEGLEEDNITERLPENSPIQ-YQENAGLGASDPKPYMKDVONQRY 590
DB 569 EKYASL-----PDNI--RLSLNKIRQGLQNNYTRDASSINEYAYLDLASGK-- 613
QY 591 LAKSOIQELIKAKOYTKLAKLLSNRHTYNISLRKLEQLFDVNP--IPSSRDIEKAKFVL 648
DB 614 ----EIDQLIKNSKFDQKQLKLSDRSRYGYQFNNOQILKALTKQKVLPSVDF-ATKSKR 668
QY 649 DKTERKNYQIYSSASPVFQNKWSLFGYRYLLGLDPRQTIHELKLGOKAGLQFEGYEN 708
DB 669 SNLEQETNYKILNFNDYFTNDFDVARFY-----TKLAHQGAET 707
QY 709 LPSDF-----NLEDLKNRIKTP--LPSQ-----KONFKLSLLDNFNNYDEIKAPE 753
DB 708 VANYPIELLKANNLID--SSVEISNPKQLFQQLRQISFKDGAKFOYFSFNNYQ---RYSN 763
QY 754 FGL--PLFLPKELRNSN-----SGGSQNSNSPWEQIISQDKONLSNQDQLAQF 803
DB 764 LGFISTLSLQGLRONVAEILDLAIVDGTQVQSPGKTQAVIPTYSTAKIA-- 815
QY 804 STKIWEKIIGDNEFDQNNRLQYKLLKLOBSWINKTRDNLNWTYLGDKLVKPKNNLEA 863
DB 816 SAVDTNEIINFIRYAQN-----ADPDE-----LSDQL-----NSSA 848
QY 864 KFRQISNQLLEIATY-----TSALSNWNYQ----- 892
DB 849 GV-EIANLGLDLVAFYKIFSQQFTDILLPISSSLTYNTHFSQITVDELEAKKAAEAAA 907
QY 893 -----DS-----GAKSTIIFEEIAELDPKVKKEKVGADVQVKPHYA 928
DB 908 EKIIDAIKRTNTSSTGIDSEFVDIDGSKNALIIQEVL---KEKDFVKKHVRLLOQNY 963
QY 929 IGF-DDNAGKFNQEVIRSS-SRTIYLKTS-GSKLEADTIDQLNQAVKNAFLGLOSFYLD 985
DB 964 LGFKDEKTNVNTFISDYNLLIFVKESDGG--QFNTSGQLDQVWIGIPTYKNVYLS 1020
QY 986 TERFGVFOKLATSLAVQHOKKEKTLPKKLNNDGYTLIHD----- 1024
DB 1021 EEE---YNNKLDTV-----KEIEDRANKGTITDVKDIESIVGTSLVVDYLKAINPNYN 1069
QY 1025 -KLKKVPVLPQISSPEKDWFEK---LNQNGSQONVNVSTFG-SITESPYF----- 1070
DB 1070 LVAKRIVLKDGELOQDNGLEGATSILVDSTSPFTNKQKILSLYLTDSPFEDSKIDNQ 1129
QY 1071 ---STNF-----QEDADLD-----QDQDSDSRQGNNSL 1095
DB 1130 LFLNTYSSIPLKIHHKAEKFKLPVNETAQEQTQNTQSSPNTQGGTSGNGNGTST 1189
QY 1096 DNOEAG 1101
DB 1190 NQQSAG 1195
```

RESULT 7
Q8G1Z2 PRELIMINARY; PRT; 1427 AA.
.ID Q8G1Z2

```
AC Q8G1Z2;  
DT 01-MAR-2003 (TrEMBLrel. 23, Created)  
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)  
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
DE Lipoprotein.  
GN Name=Ipps;  
OS Mycoplasma conjunctivae.  
OC Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.  
OX NCBI_TaxID=45361;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=22562633; PubMed=12676664;  
RX DOI=10.1128/JAEM.69.4.1913-1919.2003;  
RA Belloy L., Janovsky M., Vilei E.M., Pilo P., Giacometti M., Frey J.;  
RT "Molecular epidemiology of Mycoplasma conjunctivae in Caprinae:  
transmission across species in natural outbreaks.";  
RL Appl. Environ. Microbiol. 69:1913-1919(2003).  
DR EMBL; AJ514404; CA055813.1; -.  
KW Lipoprotein.  
SQ SEQUENCE 1427 AA; 155548 MW; 9155221EC1A91B7C CRC64;  
  
Query Match 5.2%; Score 502.5; DB 2; Length 1427;  
Best Local Similarity 20.9%; Pred. No. 2.6e-12;  
Matches 283; Conservative 185; Mismatches 478; Indels 407; Gaps 48;  
  
QY 1 MNKKSTLLATAAAIIGSTVFTVGLASKVKYRGVNPQTQGVISQGLDLSVAFKPSIA 60  
DB 1 MENKK--LIIGLGAAGAGVALFAIPVGIASTTKYDG-DPLKNVEEKVNSISNIVKDESI 57  
QY 61 NFTSDYQSVKALL-NG-KTFDPKSEFTDFVSKFDLTNNGRTVLEIPKY---QVVIS 115  
DB 58 GVNTTYASLSKLVENGHKASDVTAFGFFDF-----FTTKNQPTLITDSEWNCCEIVIK 112  
QY 116 EFSPEDDKERFRLGKLEKLEDGNTAQSATKFTYLLPLDMPKAAALQYGYIVDKNFNNL 175  
DB 113 DIADDDQDQFKIYYVQSKLDNNDIAKSNLRVKTLSFGFLPESLASPAHKGOQLFKSI 172  
QY 176 IIHPLSNFSA-----QSIKPLALTRSDFTAKLNQFNQNDLWVYLFKFFDLEAKA 227  
DB 173 RPTYGFIATLKGAEKNSVPLELTSSNDFAKDNVKAQSSDEAQQINKYFDLQSVID 232  
QY 228 NIRLOTADFSEKGNLVDPFVYSFIRNPQKQEWASDLNODQKTVRLYLITEFSPOAKTI 287  
DB 233 RIASDANNVFDNNGVTISKYKSLIKNPNTNSETVVPVLDNKOEWLFLQTEFSDQFKE 292  
QY 288 L-----KDYKYKDETFSSIDLKASNGTSLFANENDLKDQLDVDLDDVSDFGQSEETIT 342  
DB 293 ITNFDGKDFKHIDILRLNLKDLTS-----TKLEDKISITPFSVQDFYRRDKTNP 343  
QY 343 SNSQVKVPASERSLKDRVK---FKDQOKPRIEKFSLYEYDALSFYQLOELVSKNSI 399  
DB 344 FTIATKEQNANILNEFVKGXIAKADFNKPGTTRVRSPLISKAISAI---TQSVEDTSVP 403  
QY 400 KD-----LVNATLARNRFSI-----GKYNF----- 420  
DB 404 QNTGAGAGTTTTTTTADATLSOPTLAASDISALANQVRTAKDFRLVDVDFIEFTQFNGS 463  
QY 421 LFDLASHLDYFVLVSKAKIKQSSITKKLFIE-----LPI-----KISLSS----- 462  
DB 464 LFTNKEDRDQFLFVVOGLNQ-DLTKYIRFETGLENPLGSDVDFKFLNFITLDSQLD 522  
QY 463 ILGDQEPNIK--TLFKEVTFKLDNPRDVEIEKAFGLLYPGVNEEUEQARKAQRASFEK 519  
DB 523 VLYDGRPVVKGTVIAK-----SNAGTIIFTKEFSTILGAFKNSLFTISKERYASL-- 574  
QY 520 ESKSKGLKEFSQKEENSKAINNOEGLEEDNITERLPENSPIQ-YQENAGLGASDPK 578  
DB 575 -----PDNI--RLSLNKIRQGLQNNYTRDASSINE 603  
QY 579 YMIKDVONORYLAKSQIOELIKAKDYTKLAKLLSNRHTYNISLRLEQLFDVNP--IP 636  
DB 604 YAYLDLASGK-----EIDQLIKNSKFDQKQLKLSDRSRYGYQFNNOQILKALTKQKFLVP 657
```

QY 637 SSRDIEKAFVLDKTEKNKYWISASPVFONKSLFGYRYLLGLDPKQTIHELVLKLG 696
DB 658 SVEDEF-ATKSRSSLEQETNYKILFNNDYFTNDFVARFY----- 697
QY 697 QKAGLOFEGVENLPSPF-----NLEDLKNIRIKTP--LFSQ-----KDNFKLSLLDF 741
DB 698 --TKLAHQGAETVANYFIELLKANNLID--SSVEISNPKOLFQOLQISFKDNAKQYFSP 754
QY 742 NNYDGEIKAPEGL--PLFLPKELRNS-----SNSGSGNSNSPW 781
DB 755 NNNYQ---RYSNLGFTSLSLPEGLKAVEKILKFAIVDSSTTQVSPGQTPSIPT 811
QY 782 EQEIIISQFQDNLSNOQQAQFSTKIWEKIIIGDNEFDQNNRLOYKLLKDLQESWINKTR 841
DB 812 PTAKIASAVDTN-----EINFIIDYQANN----- 836
QY 842 DNLYWTYLGDKLVKPKNNLEAKFR-----QISNLOELLTAFY-----TSA 882
DB 837 -----VDPDAELSDQNSSADVEIANUGDLIVAFYKIFPSQQFTDTLLPIS 883
QY 883 ALSNNMNY-----YDGSQAK-----STIIPBEIAELDP----- 910
DB 884 SLTYNTHFSQITVDELEAEKAAEAAKILQDAIKKAMTNGTIAAQLADVDDSQNDFID 943
QY 911 ---KVKKEGVADYVQLKFHVAIGFDNAGKFNQOEVRSSRTIYLYKTSKSKLEADTIDQ 967
DB 944 GFLSEKDFVKHRIQLQYNSLGFQDEKNTENNIFISDNYNLFIFVRENKGQFNTSGQ 1003
QY 968 LNAQVKNAPLGLSFYLDTERFGVFOKLATSLAVQHKQEKTLPKLNNDGYTLIHDCLK 1027
DB 1004 LDQVIGTIFTYKNNVLAEE---YNKLDFTV---KEIEEKANTGITDVKD--- 1049
QY 1028 KPVIPOISSPEKDWF-----EGKLNQNGSQNNVNTFGSIIESP-- 1068
DB 1050 ---IESIVGTSVLDYLKAINPNVNLVAKRIVLKDGL-QTGDSGLEGASSSILVDSSTP 1105
QY 1069 -----YPTSTP----- 1075
DB 1106 FANQONKILSLYLTNDPKDKSIDNQLFDLTNYSSIPLKIIITHKAEKFKLPVNEVLAQ 1165
QY 1076 EDADLD-----QDQDSDRQGNNSLDNOEAG 1101
DB 1166 EQTDQSTQNSPNTQGGTQSGNNGNNTSGOSSG 1198

RESULT 8

Q9KGX3 PRELIMINARY; PRT; 1296 AA.
AC Q9KGX3;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Adhesin like-protein p146.
GN Name=p146;
OS Mycoplasma hyopneumoniae.
OC Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.
OX NCBI_TaxID=2099;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BQ14;
RA Verdin E., Saillard C.;
RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF27908; AAF91425.1;
SQ SEQUENCE 1296 AA; 146344 MW; 024ECB30C12D9670 CRC64;

Query Match 4.9%; Score 473.5; DB 2; Length 1296;
Best Local Similarity 21.6%; Pred. No. 3.6e-11;
Matches 309; Conservative 195; Mismatches 486; Indels 443; Gaps 65;
QY 2 KNKSTLLATAAIIIGSTVGTGVVGLASKVKYGVNPTQGVISQLGLIDSVAFKPSIAN 61
DB 3 KNNKNSLLV-TATAIVGAVFATTVGLVTRIRYKGENPRAELESLSVKQNVAFKSDVFD 61

QY 62 FTSDYQSVKALLNGKTFDPKS-----SEFTF-----VSKFDLFTNGRTVLIEPKKY 110
DB 62 NSTTYKQI-KALLFDETKLPGIDLNFISFYTAVNSKIQKF-----VTAPNKP 112
QY 111 QVVISFSPEDDKERFLGFLHKEKLEDGNIQAQATKFIYLLPLDMPKAALQGVSYIVDK 170
DB 113 PFEFINLPDDKNQNTFTLQFRAKHQLDNNYTAYSS-----ILSKIIAQAQSOFA-LADF 166
QY 171 NFN-----NL-----IIHPLSNFSAQSIKPLALTRSSDFTAKLNQFNODE 211
DB 167 NANHRKITKSFQTNIQNLRETDFSVDFSSQTSLASQKIPFLTRVEDFAADINKSGNOBE 226
QY 212 LWVYLEKFP----- 228
DB 227 AISRISKYFPDFQRYIHELKDDPNVLPFKGKIFDFITRRAGTNDTISLSANSEPSPL 286
QY 229 --IRLQADAFSFEKGNL-----VDPFVVSFIRN-----PONKEWASDL 265
DB 287 IKARL-TNEAKFELRGLNIEBAEMLEETKLVPDQFVVNLETDLKPGQAPEKSQKPSQ 345
QY 266 NQDQTVRLYLRTFES--PQAKTILKDYKIDETFLSS-IDLKASNGTSLFAN-----BN 317
DB 346 TETKTYFAEIDKILSKITMRKQLSDPKVAPQTSSSQPKQVKAS--VSAMSNLDOAQN 403
QY 318 DLKDQDLDVLDVSYFGQSETITTSNQKVPVPAERSLSLDRVKF-----KKDQKQPR 371
DB 404 KILVFNQSSNPQQQQPQSQPQSQPQSQPQSQPQSQPQSQPQSQPQSQPQSQPQ 463
QY 372 I----EKFSLEYDALSFSYQELVSPKNGIK-----DLVNAT-LARNLRFSLG--KY 418
DB 464 TDPDSKFKIRTKARDFLKEPNKTFYRSNKLKSQKLEKINSEYLSNKGIDGLGVKKY 523
QY 419 NFLFDLASH--LDYFVLVSQAKIK--QSSIFKFLFIELPIKISLSSILGQ----- 467
DB 524 -----ISNNOGIEYTFDIANAKIRDAQDGITS--HIETPVTISLWSSPFGSDNVLLKS 575
QY 468 --EPNITLFEKEVTFKLDNFR-----DVEIEKAFGLLY-----PGVN-----BE 505
DB 576 KTETFIIPYFQETTESKDKQVGHQTKELDNLQKLYVQLSELPGTSTOGSSGSSQTQEQ 635
QY 506 LEQAKAQRASF-----EKEKSKGLKEFSQKKEEN-----SKAINNOGEEEDDNIATERLP 557
DB 636 IKEVKLPTLTATFISQLEALIDGKNLASQTSQAVSVSQEVKTTFEQQAEANSTSSP 695
QY 558 EN-SPIQOQENAGLQSPDKPMKD-----VONQRYLAKSQIQELIKAKDYTKLAKLS 613
DB 696 TSPSP-----SPSPTSPKLDENIGVFNPR-----FEEIKKIS 729
QY 614 NRHTYNISLRLKEOLFV-----NPRIPSSRDIEKAFVLDKTEKNKYWISASPVFONK 670
DB 730 SBTYKYNFRANEALLDAMWGQNPPSLKDI--SQRSQRLAKDYKLVNLSKNKFLKED 787
QY 671 WSLFGYRYLLGLDPKQTIHELVLKQKAGLOFEGYENLPDFNLBDLKN-----RI 723
DB 788 YDVLAFYANLVQKDPREVLYLFEIATA-----NNLIGPEEKLDLAIEDGIFRRA 839
QY 724 KTFPLFSQDNFKLSL--LDFNNYDGEIKAPFPG--LPLFLPKELR-----RNSNSGG 773
DB 840 KATKLDKSNNGGIYGFSPNNQF---LKFHERGWMSTLYLPNEAKTKLADYQNLSSQGI 896
QY 774 SQ-----NSNSPWEQEIISQFKONLSNDQLAQFSTKIWEKIIIGDNEFDQNNRLOY 826
DB 897 SDTKIFSELNKIQPLDNIKAQSSSDSKSDSSDS-----SDAKTFTNQDILLS 946
QY 827 KL--LKDLOESWINKTRDNLTYT-----YLG-----DKLVKPKNNLEAKF-- 865
DB 947 KLTSLKSOEIAIVKK-----YETESKKYLGTENNNGSSSGTQKGSIIPEEN--KKFIL 999
QY 866 ---ROISNIOELLTATYTSAAALSN--NNNYQDSGAKSTIIFEEIA----- 906
DB 1000 ENTAKLDNLADLLAFYQAKRLNFTWSQLOQDDEDLDYQIQFEKEANNTSSSSSSSS 1059
QY 907 -----ELDPKVEKVGADYVQLKHYAIGFDPDNAGKFNQEVIRSSRTI---YLKTSG 956

```
Db 1060 SSSSSSTDTNKP--NAVEYKLYYYKI-----YNTTKVYVYTPKTIKLYLASSN 1112
Qy 957 KSKLEADITDOLNOAVKAPLQSGFYLDTERFGVFOKLATSLAVQHKQKKTLPKLLN 1016
Db 1113 TGVKQKRELMNKLVLSPAYSIFYL-----KQSE----- 1144
Qy 1017 DGYTLIHDKLKPVIPQISSPEKDFEGLKLNQSQSNVNVSTFGSIIESPYSTNFQE 1076
Db 1145 -----WDQVKITNNG----- 1155
Qy 1077 DADLDQGDSDRRGNNSLDNQEAGLLKQKLAILLGNQFIQYQONDKEIFEINLV--E 1134
Db 1156 -----QMGOTSSQGFESLEPFK-----KIQEIVHRNNKDYLVKVVTRDD 1196
Qy 1135 KVSELSFVFEKPLAKTLDNG--KTIRVLSDETMSLIVNTIETKTPMSAVPE 1185
Db 1197 AYAENAKIVHLVRVREEQGRAKTEREVEKEQ-----KETSSQVFPQ 1241

RESULT 9
Q6R5H9 PRELIMINARY; PRT; 1078 AA.
AC Q6R5H9;
DT 05-JUL-2004 (TrEMBLrel. 27, Created)
DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE Adhesin.
OS Mycoplasma hyopneumoniae.
OC Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.
OX NCBI_TaxID=2099;
RN [1]
SEQUENCE FROM N.A.
RA Bojaveit A., Cheikh Saad Bouh K., Khevar A., Shareck F., Dea S.;
RL Submitted (DEC-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY512903; AAS00638.1; -.
DR InterPro; IPR000719; Prot_Kinase.
DR PROSITE; PS00107; PROTEIN KINASE ATP; UNKNOWN 1.
SQ SEQUENCE 1078 AA; 122102 MW; 717A6B6AB6D4F45F CRC64;

Query Match 4.2%; Score 410.5; DB 2; Length 1078;
Best Local Similarity 20.6%; Pred. No. 1.1e-08;
Matches 259; Conservative 211; Mismatches 48; Indels 313; Gaps 59;

Qy 1 MNKKSTILLATAAIIIGSTVFGTVGLASVKYKRGVNPQTQGVISQLGLIDSVAFKPSIA 60
Db 1 MSKSKTKFKIGTAGIVGLGVGLTVGLSSLAKEYSESPRKIANDFAAKVSTLAFSPYAF 60
Qy 61 NFTSDYQSVKALLNGKTFDPKSGSEFTDFVSKFDFLTNNGRTVLEI-----PKKYQV 112
Db 61 ETDSYKIVKRWLVDSNN-NIRNKE--KVIDSFSTFKNGQLEKINFQDPPEYTKAKITF 117
Qy 113 VISEPSDEDDERFELGPHLKEKLEDGNIAGSATKFIYLLPLDMPKAA---LGOYSYIVD 169
Db 118 EILEIIPDVNQNFVKVFKQALQKLGNDIAKSD---IYEQTVAFAQSNLLVAFESFLK 174
Qy 170 KNFNNL-----IIHPLSNFS---AQSIKPLALTRSSDFIAKLQFNQNNQDELWYLEKFF 220
Db 175 KITEKLNQNIENLSKPTINFADKETSQKQPSLRAIDFYDLDNTARNPEDLDIKLANYF 234
Qy 221 DLEALKANI-RLQTADEFKGNLVDPPVYGFIRNPQNKQEWASDLNQDQKTVLYLRT 279
Db 235 PV-LKLNILNNAPENKLPNNLGNIFKFSFAKD-SSTNQYVSITQNG---IPSLFLKAD 288
Qy 280 FSPQAKTILKDYKYDET--FLSSIDLKASGTSLFANENLDKQDLVDLLDVSDYFGQ 337
Db 289 LSQSAEILAS---PDEVQPVNIURLMKKNSSYFLNFEDFVNNL----- 331
Qy 338 SETTITSNQKVPVAPASRLKDRVKFKKQDQKPRIKFSLYEY-----DALSFY 386
Db 332 -----TLKN---MQKEDLNAGQNLSAYEFLADIKSGFFPGDKRSSH 370
Qy 387 --SQQLVSKPNSIKDLVNATLARNLRFSLGKYNFLFDDL--ASHLDYYFLVSKAKIKQ 442
```

```
Db 371 TKSEISNLLNKENIYD-----FGKYNGFENDRLNSPNLEYSLSDAASALDK 417
Qy 443 SSITTKFLIELPIKISLSSILGDQ--BPNIK-TLFEKEVTFKLDNFRDVEIEKAFGLLYP 500
Db 418 KD---KSVIUIPYRLEIKDFADDLYPDKNILVKEGILKLTGFK-----KQPKDLP 469
Qy 501 GVNEELQARKAORASPEKSKGLKEFSQOKEENSKAINNOGLEEDDNITERLPENS 560
Db 470 NINQOI-----PKTEY---LPFPEKKEEQAK----- 493
Qy 561 PIQYQENAGLGASPDYPYMKOVQORY--YLAQSOIQELIKAKQYTKLAKLSNRHTY 618
Db 494 -LDY-----GNILNPSYNTLAKAVEALFKGNKQEIQAALDGNAY 535
Qy 619 NISLRLEQLFDVNPRI--PSSRDIEK-----AKFVLDKTEKNKYQIYSS--A 663
Db 536 EFG-AFKSVLNSWTGKIQHPEKADIQRTHLQVGLKGSVNLNQPOTTKE-QVSSLSK 593
Qy 664 SPVFQNKWSLFGYRYLLGLDPKQTIHELKVLGQKAGLOFEGYENLPSDFNLEDLKNI-- 721
Db 594 NNFFNGHQVASYFQDILLTKDLTVLSTLYDLAKWGLE-TNRAQFPKEV-FQDTKDIFA 651
Qy 722 ---RIKTPFSQKDNF-----KLSLLDFNNYDGE-IPAFEGCLPLPLKELRRNSSGG 773
Db 652 EADKLKLEWKEKDPYNQIKEIHQLSFNILARNDVIKSDGFYGVLLLPQSVKTELEGN- 710
Qy 774 SQNSNSPWEQEIISQFQDONLSNQDLAQFSTKIWEKIIIGDENEFQD--NNRLOVKKLKL 832
Db 711 -----EAQIPEALKKYSLIEN---SAFKTILDKNLLEGDTFKTFGDKAFLKAA 759
Qy 833 Q-----ESWINKTRDNLWYTY---LGDKLKVKPKQNNLEAKFRQISN-LQELLTAFYTSAA 883
Db 760 QFNFPWA-KLDDNLQYSPEAIKKGTTKEGKREEDVKVKELDNKIKGILP--QPPAA 816
Qy 884 LSNWNYYQSDGAKSTIIPEEIAE--LDPKVEKVGCA----- 918
Db 817 -----KPEAAKPVAAKPEAAKPVAAKPEAAKPEAAKPEAAKPEAAKPEAAKPEAAK 868
Qy 919 ---DVYQLKFHVAIGFDDNAGKFNQSVIRSSRTIYKTSKGSKLEAD-TIDOLNOAVK 973
Db 869 KPREDYPMAFSYKLEYTDE----NKLSLKTPELNVLELVHQSEYEEOKIKELDKTVL 924
Qy 974 NAPLGLQSYLDTERFGVFOKLATSLAVQHKQKKTLPKLNNDGYTLIHDKLKPKVIPP 1033
Db 925 NLQYQFQEVKVASDQ---YQKLSPMMTEGSSNQKKAEGAPNQ-----KKAEGAPS 974
Qy 1034 ISSPEKDWFEGLKLNQSQSNVNVST-----FGSIIESPYSTNFQEDADLD 1081
Db 975 QGKKAEGAPSGK-KAEGTSNQQTITTELNYLPDLGKKIDEIHKQ--GKNWKTVELI 1031
Qy 1082 QDG-----QDSDRQGNNSLDNQEAGLLKQKLAILLGNQFIQYQONDKEIE 1127
Db 1032 EDNIAGDAKLLYFILRDDSKSG-----DPKSSSLKVIITVKQSN-----NNQELK 1076

RESULT 10
Q6R5I8 PRELIMINARY; PRT; 1037 AA.
AC Q6R5I8;
DT 05-JUL-2004 (TrEMBLrel. 27, Created)
DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE Adhesin.
OS Mycoplasma hyopneumoniae.
OC Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.
OX NCBI_TaxID=2099;
RN [1]
SEQUENCE FROM N.A.
RA Bojaveit A., Cheikh Saad Bouh K., Khevar A., Shareck F., Dea S.;
RL Submitted (DEC-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY512894; AAS00629.1; -.
DR InterPro; IPR000719; Prot_Kinase.
```


816	-----KPEAAKPVAAKPEAAKPVATNTNTNTGSLTNKPKB-----DYFPMAFSTKLBY	864
932	DDNAGKFNQEVIRSSRTIYLTKTSCKSLEAD-TIDQLNOAVKNAPLGLQSFYLDTERFG	990
865	TDE-----NKLSTPEINVFLELVHQSEYEHQKIIEKLDKTVNLQYQFQEVKVTSEQY-	919
991	VFOKLATSLAVOHQKQKTLPKKKLNDGVTLTHDKKPKVPIQIOWSSPKDFEGKLNQ-	1049
920	-----QKLSHPMTTEGSSNOGKK-AEGAPNOG	945
1050	-----NGSQNVNVSTP-----GSIIESPYSTNFQSDADLDQDG-----	1084
946	KKAEGASNOQSTTETLNTVPLDGLKKIDELIKQ--GKWKTEVELIEDNIAGDAKLIFY	1003
1085	--QDSSROGNSLNDQEAELLKQKLAILLGNQFIQYQONDKEIE	1127
1004	ILRDDSKSG-----DPKXSLVKVITVKQSN-----NNQELK	1035
RESULT 11		
Q8I525	PRELIMINARY; PRT; 5767 AA.	
Q8I525	AC	
DT	01-MAR-2003 (TrEMBLrel. 23, Created)	
DT	01-MAR-2003 (TrEMBLrel. 23, Last sequence update)	
DT	01-OCT-2003 (TrEMBLrel. 25, Last annotation update)	
DE	Hypothetical protein.	
GN	ORFNames=PFL1930w;	
OS	Plasmodium falciparum (isolate 3D7).	
OC	Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.	
OX	NCBI_taxid=36329;	
RP	[1]	
RP	SEQUENCE FROM N.A.	
RX	MEDLINE=22255705; PubMed=12368864; DOI=i0.1038/nature01097;	
RA	Gardner M.J., Hall N., Fung E., White O., Berriman M., Hyman R.W.,	
RA	Carlton J.M., Pain A., Nelson K.E., Bowman S., Paulsen I.T., James K.,	
RA	Eisen J.A., Rutherford K., Salzberg S.L., Craig A., Kyes S.,	
RA	Chan M.S., Nene V., Shallow S.J., Suh B., Peterson J., Angiuoli S.,	
RA	Pertea M., Allen J., Selengut J., Haft D., Mather M.W., Vaidya A.B.,	
RA	Martin D.M., Fairlamb A.H., Fraunholz M.J., Roos D.S., Ralph S.A.,	
RA	McFadden G.I., Cummings L.M., Subramanian G.M., Mungall C.,	
RA	Venter J.C., Carucci D.J., Hoffman S.L., Newbold C., Davis R.W.,	
RA	Praser C.M., Barrall B.;	
RT	"Genome sequence of the human malaria parasite Plasmodium	
RT	falciparum.";	
RL	Nature 419:498-511 (2002).	
RN	[2]	
RP	SEQUENCE FROM N.A.	
RP	Hyman R.W., Fung E., Conway A., Kurdi O., Mao J., Miranda M.,	
RA	Nakao B., Rowley D., Tamaki T., Wang F., Davis R.W.;	
RL	Submitted (JAN-2003) to the EMBL/GenBank/DBJ databases.	
DR	EMBL; AE014850; AAN36472.1; -.	
DR	HSSP; P40136; 1PK0.	
KW	Hypothetical protein.	
QY	SEQUENCE 5767 AA; 691409 MW; E84A340441C67160 CRC64;	
Query Match		
Best Local Similarity 4.2%; Score 405; DB 2; Length 5767;		
Matches 406; Conservative 320; Mismatches 667; Indels 680; Gaps 104;		
QY	65 DYOSVRK-----ALINGKTFDPKSEF-----TDFVSKFDTLNNGRVLPBPKYQVVIS	115
DB	3290 EYDSLKKFKIDKVKNEYIDILDSQFEQHKDIYEKFGKNNNSQNELSNINK-DVILP	3348
QY	116 EFSPEDDKFRIGFHLK--EKLEDGNI-AQSATKFIYLLPLDMPKALGOYSIVDK--	170
DB	3349 SLISFARCDIMKLSHDVNSIELLDLLNNKELNVYISLTKTDI-ENMLDHNHYNDKND	3407
QY	171 ----NFNNLIHPLNFSQAQSIKPLATRSSDFIAKLNQ-----FNNQDELWVYL-E	217
DB	3408 HNNYNYNYFNHG-SGFIYPEDNNISKINKODLIEKLKQDSEIKYEYENKLDNNNIYVK	3466
QY	218 KFTDLEALKANIRLQTADFSEKGNLVDPFVYGFIRNPQNKQEWASDLNQDQKTVLYLR	277

816	-----KPEAAKPVAAKPEAAKPVATNTNTGSLTNKPKB-----DYFPMAFSTKLBY	864
932	DDNAGKFNQEVIRSSRTIYLTGSKSLEAD-TIDQLNOAVKNAPGLQSFYLDTERFG	990
865	TDE-----NKLSTPEINVFLELVHQSEYEHQKIKELDKTVNLQYQFQEVKVTSEQY-	919
991	VFQKLTSLAVOHQKQKTLPKKLNNDGYTLTHDKKPKVPIQIOWSSPKDFEGKLNQ-	1049
920	-----QKLSHPMTTEGSSNOGKK-AEGAPNOG	945
1050	-----NGSQNVNVSTP-----GSIIESPYSTNFQSDADLDQDG-----	1084
946	KKAEGASNOQSTTETLNTYPLDGLKKIDELIKQ--GKWKTEVELIEDNIAGDAKLYF	1003
1085	--QDSSRQGNNSLDNQEAGLLKQKLAILLGNQFIQYYOQNDKEIE	1127
1004	ILRDDSKSG-----DPKXSLVKVITVKQSN-----NNQELK	1035
RESULT 11		
Q8I525	PRELIMINARY; PRT; 5767 AA.	
Q8I525	AC	
DT	01-MAR-2003 (TrEMBLrel. 23, Created)	
DT	01-MAR-2003 (TrEMBLrel. 23, Last sequence update)	
DT	01-OCT-2003 (TrEMBLrel. 25, Last annotation update)	
DE	Hypothetical protein.	
GN	ORFNames=PFL1930w;	
OS	Plasmodium falciparum (isolate 3D7).	
OC	Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.	
OX	NCBI_taxid=36329;	
RP	[1]	
RP	SEQUENCE FROM N.A.	
RX	MEDLINE=22255705; PubMed=12368864; DOI=i0.1038/nature01097;	
RA	Gardner M.J., Hall N., Fung E., White O., Berriman M., Hyman R.W.,	
RA	Carlton J.M., Pain A., Nelson K.E., Bowman S., Paulsen I.T., James K.,	
RA	Eisen J.A., Rutherford K., Salzberg S.L., Craig A., Kyes S.,	
RA	Chan M.S., Nene V., Shallow S.J., Suh B., Peterson J., Angiuoli S.,	
RA	Pertea M., Allen J., Selengut J., Haft D., Mather M.W., Vaidya A.B.,	
RA	Martin D.M., Fairlamb A.H., Fraunholz M.J., Roos D.S., Ralph S.A.,	
RA	McFadden G.I., Cummings L.M., Subramanian G.M., Mungall C.,	
RA	Venter J.C., Carucci D.J., Hoffman S.L., Newbold C., Davis R.W.,	
RA	Praser C.M., Barrall B.;	
RT	"Genome sequence of the human malaria parasite Plasmodium	
RT	falciparum.";	
RL	Nature 419:498-511 (2002).	
RN	[2]	
RP	SEQUENCE FROM N.A.	
RP	Hyman R.W., Fung E., Conway A., Kurdi O., Mao J., Miranda M.,	
RA	Nakao B., Rowley D., Tamaki T., Wang F., Davis R.W.;	
RL	Submitted (JAN-2003) to the EMBL/GenBank/DBJ databases.	
DR	EMBL; AE014850; AAN36472.1; -.	
DR	HSSP; P40136; 1PK0.	
KW	Hypothetical protein.	
QY	SEQUENCE 5767 AA; 691409 MW; E84A340441C67160 CRC64;	
Query Match		
Best Local Similarity 4.2%; Score 405; DB 2; Length 5767;		
Matches 406; Conservative 326; Mismatches 667; Indels 680; Gaps 104;		
QY	65 DYOSVRK-----ALINGKTFDPKSEF-----TDFVSKFDELTTNGRTVLBPKYQVVIS	115
DB	3290 EYDSLKKFKIDKVKNEYIDILDSQFEQHKDIYEKFGKNNNSQNELSNINK-DVILP	3348
QY	116 EFSPEDDKFRIGFHLK--EKLEDGNI-AQSATKFIYLLPLDMPKALGOYSIVDK--	170
DB	3349 SLISFARCDIMKLSHDVNSIELLDLLNKKELNVYISLTKTDI-ENMLDHNHYNDKND	3407
QY	171 ----NFNNLIHPLNFSQAQSIKPLATRRSSDFIAKLNQ-----FNNQDELWVYL-E	217
DB	3408 HNNYNYNYFNHG-SGFIYPEDNNISKINKODLIEKLKQDSEIKYEYENKLDNNNIYVK	3466
QY	218 KFTDLEALKANIRLQTADFSEKGNLVDPFVYGFIRNPQNOKEWASDLNQDQKTVLYLR	277


```
Db 3467 KERIEKLGK--EYGDQFSINKN-----SYISNNNDNRNSSLSEK--HLHIK 3513
Qy 278 TEPFQAATILKDYKYKDETFLSSIDLKASNGTSLPANENDLKDOLVDLLDVSDYFGGQ 337
Db 3514 BEY-----KELNEXKF-----SUNLU-----NUTKNKLD----- 3538
Qy 338 SETITNSQVFPVPAERSLDRVKFRKQOQKPRIEFSLFSDAL--SPYSQJQELVSK 395
Db 3539 -----TLK--REKDKMEK--LKELOIHDKDVIIQNLKGINRELENN 3575
Qy 396 PMSIKOLVATLARNLRSFGKYNPL--FDDLASHLDYFLVSKA-KIKQSSYFKLPFIE 452
Db 3576 LNEIKSFPNEKI-----LEKENIIEKNDLINLNEEIIIMLSIKIKNNNFIKEL--E 3626
Qy 453 LPIK-----ISLKSILGDQPNIK-----TLFEKEVTFKLONFRDVEIEKAFGLLYPCVN 503
Db 3627 LOIKNLNNEINTLMDLKDSSEIRMLNHTLEEQ-----N 3662
Qy 504 BELEQARQAORASP---EKEKSIGKLFESQKKEENSKAINNOBGLBEDDNIITRLPENS 560
Db 3663 ECVKLRNKLQNYLHISNEDKYDKSSSTYNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 3716
Qy 561 PIQVOEQNAGLQSPDKPYMTKQVONQRYYLAKSOIQELIKAKDYTKLAKLLSNR-HTYN 619
Db 3717 -TRRRQDN-----QIDDLKR-----KSVLQSNHEYKQNNELRDLNNSFSYN 3758
Qy 620 ISLRLEQOLFDPNPRIPG-SRDIEKAKFVLDTKKNYQWQYSSASPVFQNKWSLFGYR 678
Db 3759 ---KUK---FQSEKQIEDFEKANMLKFYINKIE-----ENK----- 3790
Qy 679 YLLGLDPRQTHIELVKLGQK-----AGLQFEGYENLPDFF--NL-----ED 717
Db 3791 -----KKEEXHALIEVAIKEMDKENITLQLEKKNVFINDIQNNLEKKKTLLEYLEED 3844
Qy 718 LKNRIKTPLSQKDNFKLLDPNNYD-----GEIKAPEGLPL-----FLPK 763
Db 3845 IKIVQHIIMKYVQKLN---SLKDIEQLDELNENELKTNEYFTFIKOIFISIHQNYMKEE 3901
Qy 764 LRRSSNGGSGQNSNPWEQIISQFQDNLSNODLAQFQTKIWEKIIGDENFDF--Q 820
Db 3902 LNRK-----EQMLN---CQGLIKENEVIQLEKELKELTL-DKRYDKDIK 3943
Qy 821 NNRLOYKLLDQESWINKTRDNLNLYTLGDKLVKPK--NNLEAKFRQISNQLQELTAF 878
Db 3944 NQKDELRLIDRKDSY---EDEL--NHIINELEIKREINQMNKNELKNIQ----- 3991
Qy 879 YTSAAALSNWYQDSGAKSTIIFEEIAELDPKVKYKGVADYQKLFHYAIGFDNACKP 938
Db 3992 -----NENKRVEDIKT-----EGDLKYKE-----ELKYR-----EENLKH 4025
Qy 939 NOEV-----IRSSRTIYLTSGSKLEADTIDQLNO-AVKNAPLGLOSPYL 984
Db 4026 EELLIKENYMNKELMKKEELIVRENDLIKNKEDLIDKTNELMLKREELQKDKIHI 4085
Qy 985 DTERFVGFQKLATSIAVOH--KQEKTLPKKLNDNGYHTLIHDKLKPVPQISSPEKOW 1042
Db 4086 DELKNDMLMKPEKLYNAQNDIKVETKLLK-----IVEKLKDHNIIEI----- 4128
Qy 1043 FEGKLNQNGSQNVNVSTFGSIIESPYFTNFEQEDADLDQDQDSDRQGNNSLNOBAGL 1102
Db 4129 -----FDIINRRDNNISKN-----HTSNINSIDI---KNHTTNNNNNNNNNNNN 4173
Qy 1103 LKQKLAILLGNOFIQYQONDKIEF---EINVEKVSLSFRVEF-KLAKTLEDNGKT 1157
Db 4174 SLKRLGILLGLDEIG--DENELDFHFDKENESLNDLSLEYTYTFNEFKRIKLNKNN--- 4228
Qy 1158 IRVLSDETMSLIVNTTIETKTPMSAVPFTQKWEQYDPRTPIAAKTKFVLKFKDQIPV 1217
Db 4229 --LLNRE-----KQVEIYKKSVEIEKEKEIYKIE----- 4256
Qy 1218 DSGNISDKWLASIPLVTH-CQMLRL-----SPVKTIRELGLKTEQOQOQO 1263
Db 4257 -----LEKUNELLHVEQNNRKNLDLELEKYKSEDTTHIVKSLRE-----SEELNE 4301
Qy 1264 QOQOQOQKAVRKEELETYNPKDEFNLTNPLTKAHLRTLSNL--VNNDPNYK---- 1317
Db 4302 KNNKILELOQLIETSVEINIMIDKKN-NLIEKQDYEQKIDNLSIINDYEKEIKEMN 4360
Qy 1318 -----EDLVKINBAGHQ-LAFSLRANNIKRLMNTTITFADYDPPFF 1358
Db 4361 KEKLKTRKSIEMIDGEDIKKIDLDEAQLILDLEKKN-BELNINIIELIN----- 4414
Qy 1359 YYNEDWR-SIDKYLNNKGVSSHQOQAAAGNOGSLIORLNKNIKPTFTPALIALKORN 1417
Db 4415 -SNKMDRLDVLVLSNDVNLNK-----ENILNEKKMNEDEQKYNELKTIAYQKV 4462
Qy 1418 NTLNSYSDKIIMIKPKYLVERSIGVPMSTGLDGYIGSEQTKDGTSSSSQOQGFDPQPIQ 1477
Db 4463 HEFNKLYNMLSIKMNKFDSENDMEQNYT--NDNVNMQKSRNSINNKYDINDPENDTII 4519
Qy 1478 ALGLKNTYHGKLGSLIRIFDPGNELAKIKDASNKKEEKLKSYDLFKYVNLN----- 1531
Db 4520 ALNEKN-----MIIESLNEKTAHYEENYKKEELVKN---YKMINELSNKIH 4564
Qy 1532 -----YEKKSPIK-----AKGWT 1544
Db 4565 IFEDNLDPLISLNEFIKTNFTYQNIILDIINSLYKLLSKLILVENQFYKIIISIRDWL 4624
Qy 1545 NIHPDQKEYPNPKLPENYLNVLNQP-----WKVTLYNSSDPFITNLF 1588
Db 4625 FEIIKSXSYSNN--NNNNIISNNTRESVLMSLRKKGKFNKRTASVTSIYNLDF 4681
Qy 1589 V-----EPBGSD-RSGTGLKQVIOKQV--NNY---ADWGSAYLTFWYDK 1628
Db 4682 LSENVLSAEKDEINMSFNKTLNVIDDDFNSPENICIELNLFIEEDMNNIFIKL----- 4737
Qy 1629 NIITNOPVITANIADVFIFKOVK-----ELEDNTKLIAPNITQWMPNISGS 1674
Db 4738 NILENQ-----FKNIILIIHEMKKINKKNNNEKLEKKSVKILKRLN-----SISND 4789
Qy 1675 KEKPYKPTVFFGWMENESSMNSOQ-----TPTWEKIRGEFALQAL---KSS 1719
Db 4790 LEKRTBEI-----EHLNIALNKECAEKILLNQYEDIKKTYKQNLQYNEQIILIQNE 4843
Qy 1720 FDKTRFTVLITNAPLWKYGLGFONGPNFKTDWRLVFQNDNQI-----AALRVQ 1773
Db 4844 YELKSLSDQLLLE-----KNENQIKTEKINKLNSDLNHNFTSLDSKAYQLS 4890
Qy 1774 EODRPEKSS-----EDKDK-----QKWIKFKVVIPEEM--FNSGNIRFVGMQIOGPN 1819
Db 4891 ELNNEKKKKNKILLNELEBEKEKIISHLQKDPESNVVINELKEFNEMLINKLKKEEYKKE- 4949
Qy 1820 TMLPVLINSSVIYD-----FYRGTGDSNDVANLVAP---WQVKTIIFTNNAFNVKPEF 1871
Db 4950 -----KDNVYDKNSHDNYSNNNDNNNNNNNNNNSSNRNHSNNNNNNNNNNNNKSF 5002
Qy 1872 N 1872
Db 5003 H 5003

RESULT 12
Q6RS10 ID Q6RS10 PRELIMINARY; PRT; 1063 AA.
AC Q6RS10;
DT 05-JUL-2004 (TrEMBLrel. 27, Created)
DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE Adhesin.
DS Mycoplasma hyopneumoniae.
OC Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.
RN NCBI_TaxID=2099;
RX [1]
RP SEQUENCE FROM N.A.
RA Bolavert A., Cheikh Saad Bouh K., Kheyar A., Shareck F., Dea S.;
```

RL	Submitted (DEC-2003) to the EMBL/GenBank/DBJ databases.	
DR	EMBL; AY512902; AAS00637.1; --	
DR	InterPro; IPR000719; Prot kinase.	
DR	PROSITE; PS00107; PROTEIN KINASE ATP; UNKNOWN 1.	
SQ	SEQUENCE 1063 AA; 120502 MW; C31A18060CE27F05 CRC64;	
	Query Match 4.2%; Score 404; DB 2; Length 1063;	
	Best Local Similarity 20.7%; Pred. No. 1.9e-08;	
	Matches 259; Conservative 214; Mismatches 462; Indels 318; Gaps 60;	
QY	1 MNKKSTLLATAAALIGSTVGTVGLASKVKYGVNPTQGVISQLGLDSVAFKPSIA 60	884 LSNWNYQDSGAKSTIIFERIAE---LDPKVKKEKVG- 918
DB	1 MSKSKTKFGLTAGLVGLVGLTVGLSSLAKEYSESPRKIANDEFAKVSFLASPYAF 60	816 -----KPEAAKPVAAKPEAAKPVAAKPVAAKPVAAKPVAAKPVATNTGSLTNPKPE 867
QY	61 NFTSDYQSVKALLNGKTPDPKSSFTDFVSKFDFLTNNGRVLEI 112	919 DVYQLKFHYAIGFDDNAGKFNQEVIRSSRTIYILKTSKSKLE-ADTDQLNQAVKNAPL 977
DB	61 ETDSYKIVKRWLVDSNN-NIRNKE--KVIDSFSTFKNGDQLEKINFODPEYTKAKITF 117	868 DYFPMAFSYKLEYTDE---NKLSTKTPPEINVFLELVHQSVEYDQKIIEKLDKTVLNQY 923
QY	113 VISESPEDDKERFRLGPHLKEKLEDCNIAQATKFIYLLPLDMPKAA---LGQYSYIVD 169	978 GLQSFYLDTERFGVFOKLATSLAVQHOKEXTLPKLNNDDGYTLIHDKLKFPVLPQISS 1037
DB	118 EILEIIPDVNQNFVKVFOALQKLNHNGDIKSD---IYEQTVAFAKQSNLLVAEFNFSLK 174	924 QFQEVKVTSDQ---YQKLSHPMTEGTQNOQKGEPTNOG-----KKAECA 967
QY	170 KNFNLL-----IIHPLSNFS---AQSIKPLALTSSDFIAKLQNQNNODELWVYLEKFF 220	1038 PEKDFEKLQNG---NGOSQNVNVSFT-----GSIIESPYFTNFQEDADLDQDG-- 1084
DB	175 KITEKLNQNIENLSTKITNFADEKTSQKDPSTLRAIDFQYDLNTARNPEDLDIKLANYP 234	968 PS---QSKKAEPTNQSPSTSELTNYLPDLGKIDELIKKQ--CKNWKTEVELLEDNIA 1021
QY	221 DLEALKANI-RLQATDFSEKGNLVDPPVYGFIRNPQNKQEWASDLNQDQKTVRLYRTE 279	1085 -----QDSRQGNNSLDNOEAGLLKQKLAILLGNQFIQYQQNDKEIE 1127
DB	235 PV--LKNLNLNNAENKLPNNLGNIFKFP9FAKO--SSTNQVYSIQNQ---IPSLFLKAD 288	1022 GDAKLLYFILRDDSKSG-----DPKSSSLVKVITVKQSN-----NNQELK 1061
QY	280 FSPQAKTILKDYKDEET--FLSSIDLKASNGTSLFANENDLKQDLVDLLDVSDYFEGQ 337	
DB	289 LSQSAREILLAS---PDEVQPVINILRLMKDNSSYFLNFEDFVNNL----- 331	
QY	338 SETITNSQVQVPASERSLKORVKFKQOQKPRIEKFSLEY-----DALSFY 386	
DB	332 -----TLKN---MQEDLNAGQNLSAYEFADIKSGFFPGDKRSSH 370	
	RESULT 13	
QGR514	PRELIMINARY; PRT; 1077 AA.	
AC	Q6R514; 27, Created)	
DT	05-JUL-2004 (TRENBLrel. 27, Last sequence update)	
DT	05-JUL-2004 (TRENBLrel. 27, Last annotation update)	
DE	Adheein.	
OS	Mycoplasma hyopneumoniae.	
OC	Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.	
OX	NCBI_TaxID=2099;	
RN	[1]	
RP	SEQUENCE FROM N.A.	
RA	Boisvert A., Cheikh Saad Bouh K., Kheyar A., Shareck F., Dea S.;	
RL	Submitted (DEC-2003) to the EMBL/GenBank/DBJ databases.	
DR	EMBL; AY512898; AAS00633.1; --	
DR	InterPro; IPR000719; Prot kinase.	
DR	PROSITE; PS00107; PROTEIN KINASE ATP; UNKNOWN 1.	
SQ	SEQUENCE 1077 AA; 121812 MW; 8D5B2674D1138ACA CRC64;	
	Query Match 4.1%; Score 399; DB 2; Length 1077;	
	Best Local Similarity 20.8%; Pred. No. 3.2e-08;	
	Matches 262; Conservative 206; Mismatches 475; Indels 316; Gaps 60;	
QY	1 MNKKSTLLATAAALIGSTVGTVGLASKVKYGVNPTQGVISQLGLDSVAFKPSIA 60	61 NFTSDYQSVKALLNGKTPDPKSSFTDFVSKFDFLTNNGRVLEI-----PKYQV 112
DB	1 MSKSKTKFGLTAGLVGLVGLTVGLSSLAKEYSESPRKIANDEFAKVSFLASPYAF 60	61 ETDSYKIVKRWLVDSNN-NIRNKE--KVIDSFSTFKNGDQLEKINFODPEYTKAKITF 117
QY	113 VISESPEDDKERFRLGPHLKEKLEDCNIAQATKFIYLLPLDMPKAA---LGQYSYIVD 169	113 VISESPEDDKERFRLGPHLKEKLEDCNIAQATKFIYLLPLDMPKAA---LGQYSYIVD 169
DB	118 EILEIIPDVNQNFVKVFOALQKLNHNGDIKSD---IYEQTVAFAKQSNLLVAEFNFSLK 174	118 EILEIIPDVNQNFVKVFOALQKLNHNGDIKSD---IYEQTVAFAKQSNLLVAEFNFSLK 174
QY	170 KNFNLL-----IIHPLSNFS---AQSIKPLALTSSDFIAKLQNQNNODELWVYLEKFF 220	170 KNFNLL-----IIHPLSNFS---AQSIKPLALTSSDFIAKLQNQNNODELWVYLEKFF 220
DB	175 KITEKLNQNIENLSTKITNFADEKTSQKDPSTLRAIDFQYDLNTARNPEDLDIKLANYP 234	175 KITEKLNQNIENLSTKITNFADEKTSQKDPSTLRAIDFQYDLNTARNPEDLDIKLANYP 234
QY	221 DLEALKANI-RLQATDFSEKGNLVDPPVYGFIRNPQNKQEWASDLNQDQKTVRLYRTE 279	221 DLEALKANI-RLQATDFSEKGNLVDPPVYGFIRNPQNKQEWASDLNQDQKTVRLYRTE 279
DB	235 PV--LKNLNLNNAENKLPNNLGNIFKFP9FAKO--SSTNQVYSIQNQ---IPSLFLKAD 288	235 PV--LKNLNLNNAENKLPNNLGNIFKFP9FAKO--SSTNQVYSIQNQ---IPSLFLKAD 288
QY	280 FSPQAKTILKDYKDEET--FLSSIDLKASNGTSLFANENDLKQDLVDLLDVSDYFEGQ 337	280 FSPQAKTILKDYKDEET--FLSSIDLKASNGTSLFANENDLKQDLVDLLDVSDYFEGQ 337
DB	289 LSQSAREILLAS---PDEVQPVINILRLMKDNSSYFLNFEDFVNNL----- 331	289 LSQSAREILLAS---PDEVQPVINILRLMKDNSSYFLNFEDFVNNL----- 331
QY	338 SETITNSQVQVPASERSLKORVKFKQOQKPRIEKFSLEY-----DALSFY 386	338 SETITNSQVQVPASERSLKORVKFKQOQKPRIEKFSLEY-----DALSFY 386
DB	332 -----TLKN---MQEDLNAGQNLSAYEFADIKSGFFPGDKRSSH 370	332 -----TLKN---MQEDLNAGQNLSAYEFADIKSGFFPGDKRSSH 370
QY	387 --SQLQELVSKPNSIKDLVNATLARNLRLSLGKYNFLFDL--ASHLDYVFLVSKAKIKQ 442	387 --SQLQELVSKPNSIKDLVNATLARNLRLSLGKYNFLFDL--ASHLDYVFLVSKAKIKQ 442
DB	371 TKAEISNLLNKENIYD-----FGKYGKFNDRNLSNPLEYSLDAASASLDK 417	371 TKAEISNLLNKENIYD-----FGKYGKFNDRNLSNPLEYSLDAASASLDK 417
QY	443 SSTTKGLFELPDKISLSILGQD-EPNIK-TLPEKEVTFKLDNFRDVEIEKAFGLLYP 500	443 SSTTKGLFELPDKISLSILGQD-EPNIK-TLPEKEVTFKLDNFRDVEIEKAFGLLYP 500
DB	418 KD---KSVILIPYRLIKKFPADDLPYPTKONILVKGILKLTGPK-----KGPIDLP 469	418 KD---KSVILIPYRLIKKFPADDLPYPTKONILVKGILKLTGPK-----KGPIDLP 469
QY	501 GVNEELEQARKAQRASFEKSKGLKEFSQOKEENSKAINNOGLEEDDNIETRLPENS 560	501 GVNEELEQARKAQRASFEKSKGLKEFSQOKEENSKAINNOGLEEDDNIETRLPENS 560
DB	470 NINQOI-----FKTEY-----LPFEKGEQAKL-----DYGNI- 500	470 NINQOI-----FKTEY-----LPFEKGEQAKL-----DYGNI- 500
QY	561 PIQYQENAGLGASDPKPYMIQVQNYVYLAKSQIOELIKAKDYTKLAKLNSRHTYNI 620	561 PIQYQENAGLGASDPKPYMIQVQNYVYLAKSQIOELIKAKDYTKLAKLNSRHTYNI 620
DB	501 -----NPYNTQ-----LAKVEEALPKGNQBIYQALDGNVAYEF 536	501 -----NPYNTQ-----LAKVEEALPKGNQBIYQALDGNVAYEF 536
QY	621 SLRLKEQLPDVNPRI--PSSRDIEK-----AKFVLDKTEKNKYQIYSS--ASP 665	621 SLRLKEQLPDVNPRI--PSSRDIEK-----AKFVLDKTEKNKYQIYSS--ASP 665
DB	537 G-AFKSVLNSWTGKIQHEKADIQRETHLEQVKIGNSVYLNQPTAKE--QVITSLKSN 594	537 G-AFKSVLNSWTGKIQHEKADIQRETHLEQVKIGNSVYLNQPTAKE--QVITSLKSN 594
QY	666 VFQNKSLFCYRVLGLDPKQTHLVLKQKAGL-----QF--EGYENLPDSDFNLEDL 718	666 VFQNKSLFCYRVLGLDPKQTHLVLKQKAGL-----QF--EGYENLPDSDFNLEDL 718
DB	595 FFKNGHQVASYFDQLTKDKLTVLETLYDLAKWGLNETNQAPPKFAFYTKDIPAEAD- 653	595 FFKNGHQVASYFDQLTKDKLTVLETLYDLAKWGLNETNQAPPKFAFYTKDIPAEAD- 653
QY	719 KNIRIKTFLPSQDNF----KLSLLDPNNYDGE--IKAPEFGPLFLPKELRRNSNSGG 773	719 KNIRIKTFLPSQDNF----KLSLLDPNNYDGE--IKAPEFGPLFLPKELRRNSNSGG 773
DB	654 --KLFLWKKKDPYNOIKIEIHLQSFNILARNDVIKSDGYGVGLLIPQSVKTELEGKN- 709	654 --KLFLWKKKDPYNOIKIEIHLQSFNILARNDVIKSDGYGVGLLIPQSVKTELEGKN- 709
QY	774 SQNSNSFWBEIISQFDQNLSNQDLAQSTKIWEKIIGDENEFQD--NNRLQYKLLKDL 832	774 SQNSNSFWBEIISQFDQNLSNQDLAQSTKIWEKIIGDENEFQD--NNRLQYKLLKDL 832
DB	710 -----EAQIFEALKKYSLIEN---SAFKTITLDKNLEGTDFTFFGDFLKAFFLAA 758	710 -----EAQIFEALKKYSLIEN---SAFKTITLDKNLEGTDFTFFGDFLKAFFLAA 758
QY	833 Q----ESWINKTRNLNWTY----LGDKLKVPKNNLEAKFRQITSN-LOBLLTAFVTSAA 883	833 Q----ESWINKTRNLNWTY----LGDKLKVPKNNLEAKFRQITSN-LOBLLTAFVTSAA 883
DB	759 QFNPFAPWA-KLDNQLQYSFEAIKKGTTKEGKEEYDVKVKELDNKIKGILP--QPPAA 815	759 QFNPFAPWA-KLDNQLQYSFEAIKKGTTKEGKEEYDVKVKELDNKIKGILP--QPPAA 815

```
QY 387 --SQQLVSPKNSIKDLVNLARNLRFSLGKYNFLFDDL--ASHLDYYFLVSKAKIKQ 442
DB 371 TKAEISNLLKKNENIYD-----FGKYGKENDRLNSPNLEYSLDASASLDK 417
QY 443 SSITKKLFIEPIKLSKSSILGDO-EPNIK-TLFEKEVTPKLDNFRDVEIEKAFGLLYP 500
DB 418 KD---KSVLPYRLEIKDFADDPKDNILVKEGILKJGFK-----KGPKIDLP 469
QY 501 GWNELEQAQKASFEKSKGLKFSQOKSEKSKAINNQEGLEDDNITERLSPNS 560
DB 470 NINQI-----FKTEY-----LPFEKGEQAKL-----DYGILN--FNNT 505
QY 561 PIYQOENAGLGASDPKPYMKDQVQNYRYLAKSQIOELIKAKDYTKLAKLSNRHTYNI 620
DB 506 -----QLAKVEALFKGNKQEIYQALDGNAYEYF 536
QY 621 SLRLKEQLFDVNPRI--PSSRDIEK-----AKFVLDTKTKNYQWQIYS--ASP 665
DB 537 G-AFKSVLNSWTGKIQHPKADIQRTRHLEQVKIGSNVNLNQPTTKE-QVVISLKSNN 594
QY 666 VFONKWSFGYVRYLGLDPRQTHLVKLGQKAGL-----QF-EGYENLPSDF--NLED 717
DB 595 FFKGHQVASVFDLTKDKLVLETLVDLAKKGLWTNWAQFPKGAFOYTKDIPAEADK 654
QY 718 LKNIRIKTFLFSQDNF---KLSLLDFNNYYDGE-IKAPEFGLFPLPKELRNRSSNG 772
DB 655 LKFLESK-----KKDPFNQIKIEIHQLSFNILARNDVIKSDGYGVLLLPQSVKTELEGN 709
QY 773 GSQNSNSPWEIISQFKDQNLNSQDLAQASTKIWEKIIIGDENFDO--NRLQYKLLKD 831
DB 710 -----EAQIFALKKYSLIEN---SAFKTTILDKNLLEGTDFKTFGDFLKAFFLKA 757
QY 832 LQ---ESWINKTRNLNWTY----LGDKLKVKPKNNLEAKFROISN-LOELLTAFYTS 882
DB 758 AQFNFAFWA-KLDNLQYSPFAIKKGETTGKGRKEVDKVKKELDNIKIGILP--QPPA 814
QY 883 ALSNNNYYQDSGAKSTIIFBEIAR---LDPRKVEKVGA-----918
DB 815 A-----KPEAAKPVAAKPEAAKPVAAKPEAAKPVAAKPEETTKPVATNTNTNGFSLT 866
QY 919 -----DVTQLKPHYAIGFDDNAGKPNQEVIRSSRTIYLTGSGSKLEAD--TIDQLNQAV 972
DB 867 NKPKEDYFPMAFSYKLEYTDE----NKLSTKTPINVFELVHQSXEYBEQKIIKELDKTV 922
QY 973 KNAPLGQSFYLDTERFGVFOKATSLAVOHKQKEKTLPKKLNDGYTLIHDKLLKVP 1032
DB 923 LNLQYQFEVKASDQ---YQKLSHPMTTEGSSNGKKAEGAPNQ-----KKAEGAP 972
QY 1033 QISSPSEKDMFEGKLNQNGSQNVNVT-----FGSIESPYFTNFQEDADL 1080
DB 973 SQGKKAEGAPSQK-KAEGTSNQQTTLTNYLPDLGKIDIEIKKQ--GKNWKTEVEL 1029
QY 1081 DDG-----QDSSQGNNSLDNQEAGLLKQKLAILLGNQFYQYQOQNDKETE 1127
DB 1030 TEDNIAGDAKLYLFDLDDSKSG-----DPKKSUKVITVKQSN-----NQNELK 1075
```

RESULT 14

```
Q8WR55 PRELIMINARY; PRT; 2965 AA.
AC Q8WR55
ID Q8WR55
DT 01-MAR-2002 (TReMBLrel. 20, Created)
DT 01-OCT-2003 (TReMBLrel. 25, Last sequence update)
DT 01-OCT-2003 (TReMBLrel. 25, Last annotation update)
DE Normocyte-binding protein 1.
GN Name=NBP1;
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporidia; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Malayan Camp K-;
```

```
RX MEDLINE=21590493; PubMed=11733572;
RA Rayner J.C., Vargase-Serrato E., Huber C.S., Galinski M.R.,
RA Barnwell J.W.;
RT "A Plasmodium falciparum homologue of Plasmodium vivax reticulocyte
RT binding protein (PvRBP) defines a trypanin-resistant erythrocyte
RL invasion pathway.";
RL J. Exp. Med. 194:1571-1581 (2001).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Malayan Camp K-;
RA Rayner J.C., Barnwell J.W., Galinski M.R., Tran T., Huber C.S.,
RA Corredor V.;
RL Submitted (AUG-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF411932; AAL38221.2;
DR ProDom; PD001963; Botulinum; 2.
SQ SEQUENCE 2965 AA; 356914 MW; 98B077462826A8FE CRC64;
```

Query Match 4.1%; Score 395.5; DB 2; Length 2965;
Best Local Similarity 18.7%; Pred. No. 1.7e-07;
Matches 354; Conservative 325; Mismatches 679; Indels 533; Gaps 88;

```
QY 50 IDSVAPKPSIANFT---SDYOSVKKALLNG-----KTFDPKSEFTDFVSKF 93
DB 777 IDDIKIQETLKOITHIVNNIKTKIKDKLLKSEFIQHLIKYMNERYQMGGYNLLINYINQY 836
QY 94 DFLTNNGRTVLEIPKYYQVVISPEFSDDEKRFPLGFLHKEKLEDG---NIAQSAKF 149
DB 837 EEENNMMQVITTRNIQKIYVDNIYAKEKE-IRSGQYKDFITSRKNIYNIENISQNV 895
QY 150 YLLPLDMPKALGOYSYIVDKNFNLLIHLPSNFSASIKPLALTRSDFTAKLNQNRNQ 209
DB 896 DMIKNEBK---KIQNCVDK-YNSI-----KQYVMFKNGDTQD-----ENNMMN 937
QY 210 DELWVLEKFPDLEALKANRLQATDFSEKGNLVDPPVYSFIRNPQNKQEWASDLNQD-- 268
DB 938 D---IYDKLIVPLDSIQNDIKYNTENF-----ITFNKINTHKK-----NQEM 980
QY 269 -QKTVRLYLRTFSPQAKTILKQYKDETFLLSIDLKASNGTSLFANE--NDLKQDL--- 323
DB 981 MEBFIYVYK---LKIPIK-----LNLISLKACEKNKSNITNLDKTELK 1024
QY 324 ---DVLDDVSDVFGGSEIITS-----NSQKVPVPASERSLKRVFKKQDQKP 370
DB 1025 VTIEIDLL-----QKDLTSGISNKNVLLNLDLKEI---EYIIVHKKKSN-- 1071
QY 371 RIEKFSLYEYDALSFYSQ-----LQELVSKPN---SIKDLVNATLARNLRFSLGKYNF 420
DB 1072 --DLFTYEQSKNYFYFNKKNDNFDIQTINKNNEWLAIKYINEKNYQTLVEKKINV 1129
QY 421 LFDOLASHLDYFLVSKAKYKQSSITKKLFIEPIKLSKSSILGDOEPNITKLFKEVT 480
DB 1130 LLHNSKYVQVYF---DHIINLILQKNYLENTLTKIQ-----DNEHSLYALQONEEY 1180
QY 481 FKLDNFD-----VEIEKAFGLLYPGVNELEQA-----RKAQASFEKE 520
DB 1181 QKVXNEKDKQNEIKKIQLEIKNKNIDILTYENNIEQIEQKIELKTNQAKDDQIVNTLINE 1240
QY 521 KSKKGLKEFSQOKEENSKAINN-OEG-LEEDDNITERLPENSPYQOQENAGLGASDPK 578
DB 1241 VKKLIIVYEKVDNQIISNVLKNYEEGKVEYDKNVVQNVNDADDNDIDEINDEINDID 1300
QY 579 YM-----IKDVQORYYLAQSIOELIKAKDYTKLAKLSNRHTYINSLR--LKEQLPDV 631
DB 1301 EINDIDEIKDIDHIKHFDDTKHFDDIYHADD-----TRDEYHIALSNYIITELRNI 1351
QY 632 NPR-----IPSSRDIEKAKFVLDDKTEK--NKYW-----QIYSSASP 665
DB 1352 NLQEKNNIIKIFKESAKHEIKKESQINKETKMDVNVQNRDIDRQMLDIYE--- 1408
QY 666 VFQKWSLFGYRYLLGLDPRQTHLVKLGQKAGLQFEGYENLPSPFNLEDKKNIRIKT 725
DB 1409 -LDEKYSEFN-----KTKIEENINRENINNVIEWYKNIIEYFLAHMNDQDKA 1457
```



```
QY 919 -----DVYQLKFHYAIGFDDNAGKFNQEVIRSSRTIYLTSGKSKLEAD-T 964
Db 864 TWTGFSLTNKPREDYFPMAFSYKLEYTDS---NKLSTKPTINVFLELVHQSEYEEQKI 919
QY 965 IDQLNOAVKNAPLGLQSFYLDTERFGVFOKLATSLAVQHKQEKTLPKKLNNDGYTLIHD 1024
Db 920 IKELDKTVNLNQYQFQEVKVTSEQ---YQKLSHPMTTEGSPNQGKKAEGAPNQG----- 970
QY 1025 KKKPV-----IPQISSSPKDMFEGKLNQ---NGSQNVNVSTF-----G 1062
Db 971 --KKAEGAPSGQKKAEGAPNQGKKAEGPSQKKAEGASNQSTTTLTNLYPELGKKID 1028
QY 1063 SIIEGPFSTNFOEDADLDQDQ-----QDSSRQGN-----NSLDN 1097
Db 1029 BIIKKQ--GKNWTEVELIEDNIAGDAKLLYFVLRRDSSKSGDPKKSXVKVITVKQSNNN 1086
QY 1098 QEAGLLKQK 1106
Db 1087 QE---LQSK 1092

RESULT 17
Q6R5H7
ID Q6R5H7 PRELIMINARY; PRT; 1092 AA.
AC Q6R5H7;
DT 05-JUL-2004 (TrEMBLrel. 27, Created)
DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE Adhesin
OS Mycoplasma hyopneumoniae.
OC Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.
OX NCBI_TaxID:2099;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 25934;
RA Boisvert A., Cheikh Saad Bouh K., Kheyar A., Shareck F., Dea S.;
RL Submitted (DEC-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY512905; AAS00640.1; -.
DR InterPro; IPR000719; Prot kinase.
DR PROSITE; PS00107; PROTEIN KINASE ATP; UNKNOWN 1.
SQ SEQUENCE 1092 AA; 123324 MW; 624518A7ACA3DB54 CRC64;

Query Match
Best Local Similarity 20.3%; Pred. No. 6.5e-08;
Matches 259; Conservative 209; Mismatches 473; Indels 335; Gaps 58;

QY 1 MKNKKSTLLATAAIIIGSTVGTGVLGSKVKYRGVNPQTQVISQLGLIDSVAFKPSIA 60
Db 1 MSKKSTFKITGLTAGIVGLGVGLTGLSSSLAKYRSERPKIANDPAKAVSTLAFSPYAF 60
QY 61 NPTSDYQSVKALLNGKTFDPKSEFTDFVSKPDLTNNGRTVLEI-----PKYQV 112
Db 61 ETDSYKIVKRWLVDSNN-NIRKE--KVIDSFFTKNGDQLEKINFQDPPEYTKAKITF 117
QY 113 VISEFSPDDKERFRLGHUKEKLEGNIAQATKFIYLLPLDMPKAA---LQOYIYVD 169
Db 118 EILEIIPDDVNQNFVKFQALQKHLNGDIKASD---IYEQTAVAFQSNLLVAEFNFSLK 174
QY 170 KNFNLI-----IIHPLSNFS---AQSIKPLATRSSDFIAKLNQFNODELWVLEKPF 220
Db 175 KITEKLNQQTENSTKTIENADESKTSQKDPSTLRAIDFQYDUNTAENAEIDLKLANYP 234
QY 221 DLBAKANI--RLQTADFSFEKGNLVDPVYVSTFRNPNQKEMASDLNQDQKTVRLYRTE 279
Db 235 PV--LKNLINRLNNAENKLPNNLGNIFEFSFAKD--SSTNQYVSIQNO---IPSLFLKAD 288
QY 280 FSPQAKTILDKYKKBOT--FLSIDLKASNGTSLFANENDLKQDLVDLLDVSDFEGGQ 337
Db 289 LSQSAREILAS---PDSEVQVINILRLMKXNDSYFLNFEDFYNNL----- 331
QY 338 SEITITSNQVKVPASERSLKRVPKQKQKQKPRIEKFSLEY-----DALSFY 386
```

```
Db 332 -----TLKN---MOKEDLNAGQNLSAVEFLADIKSGFPDGPGRSSH 370
QY 387 --SQQLVSVKPSIKDLVNATLARNLRFSLGKTNFLFDDL--ASHLDYYFLVSKAKIKQ 442
Db 371 TKAEISNLLNKENIYD-----FGKYNKPNDRNLNFPNLEYSIDAASASLDK 417
QY 443 SSITKCLPIELPIKISLSSILGDQ-BPNIK-TLFEKEVTFKLNFRDVEIEKAFGLYP 500
Db 418 KD---KSILIPYRLIKDKFPADLDYPTDNLIVKESILKLTGFK-----KGPIDLP 469
QY 501 GVNBELEQARQAASPEKESKGLKEFSQKENSKAINNQGLEBDDNITERLPENS 560
Db 470 NINQOI-----FKTEY---LPFFEKGEEQAKL-----DYGNI----- 500
QY 561 PIQOQENAGLGPDKPYMIKDVQNRQYILAKSIOELIKAKDYTKLAKLLSNRHTYNI 620
Db 501 -----NPNTQ-----LAKVEEALFGKNQOEIYQALDGNAYEF 536
QY 621 SLRLKEQLFDVNPRI--PSSRDIEK-----AKFVLDKTEKNKYWQIYSS--ASP 665
Db 537 G-AFKSVLNSWTGKIQHPEKADIQRFTHLEQVGLGNSVNLNPQPTKE-QVSISSKSN 594
QY 666 VFQNKWSLFGYRYLLGLDPKQTIHELKVGOKAGLQEGYENLPSDFNLEDKNIKRT 725
Db 595 FFKNGHVASVFDLLTKDKLTVLETLYLAKKWL-----ETNRAQFPKEVQYTK--- 646
QY 726 PLFSQKDNFK-----LSLLDFNNYDGE-IKAPERGLPLFLPKELRRNSS 769
Db 647 DIFAEADKLFLEGKKDPYNOIKHQLSNIARNDVIKSDGYGVLLLPQSVKTELE 706
QY 770 NSGSGSNSPWEQIIISQFKDQNLSDQLAQFSTKIWEKIIGDENEFQD--NNRLOYKL 828
Db 707 GKN-----EAQIFEAALKKYSLIEN--SAFKTTILDKNLLSGTDFKTFGDFLKAPF 754
QY 829 LKDLQ-----EGWINKTRNLNWTY-----LGDKLKVKPKNNLEAKPRQISN-LQELLTAFY 879
Db 755 LKAAQFNFPAPWA-KLDDNLQVSPFAIKKGHTTKGKEEVDKKVKELDNKIKGILP--Q 811
QY 880 TSAALSNNWNYQDSGAKSTIIFETAEID-----PKVKEKYGA----- 918
Db 812 PPAA-----KPEAAKPVAAKPEAKPETTKPVAAKPEAAKPVAAKPVAAKPVATNTN 863
QY 919 -----DVYQLKFHYAIGFDDNAGKFNQEVIRSSRTIYLTSGKSKLEAD-T 964
Db 864 TWTGFSLTNKPREDYFPMAFSYKLEYTDS---NKLSTKPTINVFLELVHQSEYEEQKI 919
QY 965 IDQLNOAVKNAPLGLQSFYLDTERFGVFOKLATSLAVQHKQEKTLPKKLNNDGYTLIHD 1024
Db 920 IKELDKTVNLNQYQFQEVKVTSEQ---YQKLSHPMTTEGSPNQGKKAEGAPNQG----- 970
QY 1025 KKKPVIPQISSSPKDMFEGKLNQ-----NGSQNVNVSTF-----GS 1063
Db 971 --KKAEGAPSGQKKAEGAPNQGKKAEGAPSDQKKAEGASNQSTTTLTNLYPELGKKIDE 1029
QY 1064 ITESPVSTNFOEDADLDQDQ-----QDSSRQGNNSLDNQEAGLLKQKLAIDL 1111
Db 1030 IIKKQ--GKNWTEVELIEDNIAGDAKLLYFVLRRDSSKSG-----DPKSSSLKVKITVQ 1082
QY 1112 GNQFIQYQOONDKEITE 1127
Db 1083 SN-----NNQELK 1090

RESULT 18
Q49542
ID Q49542 PRELIMINARY; PRT; 1108 AA.
AC Q49542;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE P97.
OS Mycoplasma hyopneumoniae.
OC Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.
```


[illegible]


```
Db 118 EILEIIPDDVNQNFVKVFOALQKLHNGDIAKSD-----IYEQTVAFAKQSNLLVAEFNFSLK 174
Qy 170 KNFNML-----IIHPLSNFS---AQSIKPLALTRSSDFIAKLQNFQNDLWVYLEKFP 220
Db 175 KITEKLNQIENLSKTIITFADEKTSQSQPSTLRAIDFQYDLNTARNAEDLDIKLANVF 234
Qy 221 DLEALKANI-RLQTADFSEKGNLVDPPVYFIRPNQOKEWASDLNQDQKTVRLYLRT 279
Db 235 PV--LKNLNLNAPENKLPNLFNIFEFSPAKD--SSTNQYVSIQNO---IPSLFLKAD 288
Qy 280 FSPQAKTILKDYKDET--FLSSIDLKASNGTSLPANENDLKQOLDVDLLDVSDVFGQ 337
Db 289 LQSAKEILAS---PDEVOPVNIILRMKONSFLNFEDFVNL-----331
Qy 338 SETITSNSQVFPASERSLDRVRFKDDQKPRIEKFSLEY-----DALSPY 386
Db 332 -----TLKN---MQKEDLNAKGQNL SAYEFLADIKSGFFPGDKRSSH 370
Qy 387 --SOLQELVSKNSIKOLVNATLARNLRFSLGKYNFLFDDL--ASHLDYFLVLSKAKIKQ 442
Db 371 TKSEISNLNKKENIYD-----FGYNGKFNDRLNSPNLEYSLSLDAASGLDK 417
Qy 443 SSITKGLFTELPKISLSKSIILGQD-EPNIK-TLPEKEVTFKLDNPRDVEIEKAFGLXP 500
Db 418 KD---KSIILIPVRLKIDKFPADDLYPTKKNILVKEGILKLTGPK-----KGPKIDLP 469
Qy 501 GVNEELEQAKORASFEKSKKGLKFEFSQOKEENSKAINNQOELEEDDNTERTLPENS 560
Db 470 NINQOI-----FKTEY---LPFFEKGKEQAKL-----DYGNIL-----500
Qy 561 PIQYOENAGLGNPKPMKDVQNRYYLAKSQIQLIKADYTKLAKLNSNRHTYNI 620
Db 501 -----NPYNTQ-----LAKVEEALFGKNKQEIYQALDGNVAYEF 536
Qy 621 SLRLKQLFDVNPRI--PSSRDTEK-----AKFVLDTKKNYKQWYSS--ASP 665
Db 537 G-AFKSVLNSWTGKIQHPEKADIQTRHLEQVKGISNVLNQPOPTKE-QVSSLSKSN 594
Qy 666 VFQNKWSLFGYRYRLGLDLPKQTIHBLVKLGQKAGL-----QF--EGYENLPSPDFNLEDL 718
Db 595 PFKNGHQVASYFDLLTKDKLTVLETLYDLAKKMGLETNQAFPKEAFQYTKDIPAEAD- 653
Qy 719 KNTRIKTPLFSQDNF-----KUSLLDPNNYDGE-IAKPEGLPLPLPELRLRNSNSG 773
Db 654 ---KLKFLWKKKDPYNQIKELHQLSFNLARNDVIKSDGFGVLLLPQSQVTELEGKN- 709
Qy 774 SQNSNSPWEQELISQFKDQNLNQDLAQFSTKIWEKIIGDENEFQD--NNRLQYKLLKDL 832
Db 710 -----EAQIPEALKKYSLIEN---SAFKTILDKNLLGCTDFTKFGDFLKAPFLKAA 758
Qy 833 Q-----BSWINKTRDNLWYTY---LGDKLKVKPKNNLEAKFROISN-LOELL-----875
Db 759 QFNFPAPMA-KLDNQLQYSEAIKKGTTKEGKREVEVDKVKKELDNKIKGILLPOPPAAKP 817
Qy 876 -----TAFYTSAAALSNNWNYQDSKAKSTIIFEEI 905
Db 818 EAAKPVAAKPEAAKPVAAKPEAAKPVAAKPEAAKPVAA-----KPEAAKPVAAKPEA 869
Qy 906 AE---LDPKVEKVG------DVYQLKFHYAIGFDDNA 935
Db 870 AKPVAAKPEAAKPVAAKPVAAKPEAAKPVATNTGSLTNPKEDYFPMAPFSYKLETYDE- 928
Qy 936 GKFNQEVIRSSRTIYLTKSGSKLEAD-TIDQLNQAQVKNAPLGLOSFYLDTERFVGQK 994
Db 929 ---NKLSLKTPENVFLVHQSEYEQKIIKELDKTVLN-----965
Qy 995 LATSLAVQHQKQKTLPKKLNNDGYTLIHDKLKPVTIPOSSEPKDWFEGKLNQ---N 1050
Db 966 -----LQYQFOE---VKVASDQY---QKLSHPMMTEGSSNQGKK-AEGAPNQGKAE 1010
Qy 1051 GQSQNVNVST-----FGSITESPYFTSNFOEDADLDQD-----QD 1086
Db 1011 GTSNQQTNTTELNYLPKELGKKIDELIKKO--GKNWKEVELIEDNAGDAKLLYFVLRD 1068
```

```
Qy 1087 DSRQGNNSLNDQEBAGLLKQKLAILLGN 1113
Db 1069 DSKSG-----DPKSSSLKVKITVQSN 1090

RESULT 20
ID Q7RFS2 PRELIMINARY; PRT: 2740 AA.
AC Q7RFS2;
DT 01-MAR-2004 (TREMBLrel. 26, Created)
DT 01-MAR-2004 (TREMBLrel. 26, Last sequence update)
DT 01-MAR-2004 (TREMBLrel. 26, Last annotation update)
DE 235 kDa thoptry protein.
GN Name=PY04630;
OS Plasmodium yoelii yoelii.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=73239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=17XNL;
RX PubMed=12368865; DOI=10.1038/nature01099;
RA Carlton J.M., Angiuoli S.V., Suh B.B., Kooij T.W., Perte A.M.,
RA Silva J.C., Ermolaeva M.D., Allen J.E., Selengut J.D., Koo H.L.,
RA Peterson J.D., Pop M., Kosack D.S., Shumway M.F., Bidwell S.L.,
RA Shallom S.J., van Aken S.E., Riedmuller S.B., Feldblyum T.V.,
RA Cho J.K., Quackenbush J., Sedegah M., Shoabi A., Cummings L.M.,
RA Florens L., Yates F.R. III, Raine J.D., Sinden R.E., Harris M.A.,
RA Cunningham D.A., Preiser P.R., Bergman L.W., Vaidya A.B.,
RA van Lin L.H., Janse C.J., Waters A.P., Smith H.O., White O.R.,
RA Salzberg S.L., Venter J.C., Fraser C.M., Hoffman S.L., Gardner M.J.,
RA Carucci D.J.;
RT "Genome sequence and comparative analysis of the model rodent malaria
RT parasite Plasmodium yoelii yoelii."
RL Nature 419:512-519(2002).
CC -! CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
DR EMBL; AABL01001422; EAA16521.1; -.
DR HSSP; P03069; 1GCM.
DR InterPro; IPR006499; ReticulocyteBP.
DR TIGRFAMs; TIGR01612; 235kDa-fam; 1.
SQ SEQUENCE 2740 AA; 321463 MW; 61A4058FF525E599 CRC64;

Query Match 4.0%; Score 390; DB 2; Length 2740;
Best Local Similarity 18.8%; Pred. No. 1.5e-07;
Matches 428; Conservative 324; Mismatches 775; Indels 748; Gaps 102;

Qy 39 PTQGVISQLGL-----IDSVAFKPSIANFTDQYQSVKALLNGKTFDPKSSBFTDFVSKF 93
Db 520 PFRSII---GFVEWRINQTLTEVEAGVKASYEVKWKRSVMEINKLEENKVIKL 576
Qy 94 DFLTN-----GRVLEIPKAYQVVISPESSPDDKFERFLGPHLKEKLED 138
Db 577 ETQINDLFNQYLKINDENIYLNKLEKELKIKNI-----SDKNEYVVKAVDLKKTEN 630
Qy 139 GNI-AQSATKFIYLPDLP-----KALAQ-YSYIVDKNFNNLIHPLSNF 183
Db 631 NNIIYDELTK---TSPYQVPEHLKNTDTTYNTIKLELSQIYEDDIDKLYNEL-----679
Qy 184 SAQSIKPLALTRSSDFIAKLQNFQNDLWVYLEKFPDLEALKANI--RLQTADFSEK 241
Db 680 -----SSIVOENIDNVED---KTKLDDLQSKIDNVYSKIQ-----712
Qy 242 NLVDPFVYFIRPNQOKEWASDLNQDQKTVRLYLRTFESPOAKTILKDYKVKDETFLSS 301
Db 713 NMENETVESHLTWIETNKNKLSD---TILAKKIYIGEISKDLNKTLEDKFNKEK-----764
Qy 302 IDLKASGTSFLANENDLKQOLDVDLLDVSDVFGQ-----SETITSNSQVXP 349
Db 765 ---ELSNKINDYAKENDQLNVYKSKISEIRNHNSQINIDNTKEGEAKQNYDKSNEHMTK 821
Qy 350 VPASERSLK---DRVFKKDKQKPRIKEF-----SLYEYDALSPYSQLQELVSKP-----396
```


"A Plasmodium falciparum homologue of Plasmodium vivax reticulocyte binding protein (PvRBP1) defines a trypsin-resistant erythrocyte invasion pathway." J. Exp. Med. 194:1571-1581(2001).

RT	SEQUENCE FROM N.A.	
RC	STRAIN=768;	
RA	Rayner J.C., Barnwell J.W., Galinski M.R., Tran T., Huber C.S.,	
RA	Corrator V.;	
RL	Submitted (AUG-2002) to the EMBL/GenBank/DBJ databases.	
DR	EMBL: AF411933; AAL38222.2; -	
DR	ProDom: PD001963; Bofulinum; 2;	
SQ	SEQUENCE 2976 AA; 358102 MW; C00P4D4A1556867C CRC64;	
	Query Match 4.0%; Score 390; DB 2; Length 2976;	
	Best Local Similarity 18.6%; Pred. No. 2.8e-07;	
	Matches 352; Conservative 325; Mismatches 684; Indels 532; Gaps 87;	
QY	50 IDSVPKPSIAFT---SDYQVKKALLG-----KTFDPKSEFTDPVSKF 93	
DB	779 IDDIKIQTQLQTHIVNNIKTKIKDLEKEFIQHLIKYNNERYQNNQGYNNLTNYINQY 838	
QY	94 DFLTNGRTVLEIPKQYQVVISFSPEDDOKFRFLGFLHLEKLEDEG---NIAQATKFI 149	
DB	839 EENNMMKQYITIRNIQIKYDNIYAKEKE-IRSGQYKDFITSRKNIYINRENISKV 897	
QY	150 YLLPDLMPKAALQYQYIYVDKPNLIIHPLSNFSAQSIKPLALTRSSDFIAKLNQFNQ 209	
DB	898 DMKNEKK---KIQNCYDK-YNSI-----KQCVKMLKNGDTQDENNNNNNNN 943	
QY	210 DELWVLEKPFDEALKANIRQATDFSEKGNLVDVFFYIRPNQOKEWASDLNQDQ 269	
DB	944 ND--IYDKLIVPLDSIKQNDIKYNTENHNF-----ITFNKINTHNKK-----NQB- 986	
QY	270 KTVRLVLRTEFSQAQITLKVYKQDETFLS--SIDLKASNGTSIFANE-NDLKQDL--- 323	
DB	987 -----MMEEFI-----YAYKELKILKILNISLAKACENKNSINTLNDKTQELKKI 1031	
QY	324 ---DVDLLDVSYFGQSEITFS-----NSQKVPVPASERSLKDRVFKDQKQRP 370	
DB	1032 VTHEIDLL-----QKDLTISOIGNVKNVLLNDLLEKEI---EQVIIDVHLKXKSN-- 1078	
QY	371 RIEKPSLEYDALSFSQ-----LOELVSKEN---SISKDLVNATLARNLFSGLKYNF 420	
DB	1079 --DLPTYYEQSKNYFYFKNKONCFDIQTKNGNEMLAKYININEKNKYQTYLEKKNV 1136	
QY	421 LPDDLASHLDYFLVSKAKIKQSSITKGLFIELPIKISLSSILGQEPNITLFEKVT 480	
DB	1137 LHNSKSYVOYFY---DHIINLILQKNVLENTLTKIQ-----DNEHSLVALQOQNEY 1187	
QY	481 FKLDNFRD-----VBEIKAFGLLYPGVNEELEQA-----RKAQASFEKE 520	
DB	1188 QKVNEKDQNEIKKIKQIEKNNDILTENNIEQIEQKNIELTKNAQNKDDQIVNTLNE 1247	
QY	521 KSKKGLKFPQSQEENSKAINN-QEG-LEEDNITELPENSPYQOQENAGLGASDPKP 578	
DB	1248 VKKKIITYTEKVDNQISNLKNEYEGKVEYDKVQNVQNDADPTNDIDEINDIDEIND 1307	
QY	579 YM-----IKDVQORYLAKSQOELIKAKDYTKLAKLLSNRHYNISLR--LKEOLFV 631	
DB	1308 EINDIDEIKDHIKHFDQTHFDIYHADD-----TRDEYHALSNYIKTELNI 1358	
QY	632 NPRISSRDIEKARFLVDKTEKNKYQWQIYSSASPVFQNKWSLFGYRYLYGLDPKQTIHE 691	
DB	1359 N-----LOEIKNNIILIKIFEKSAHKEIKESQINKKEFTKMDVWINQLRID-RQMLDL 1412	
QY	692 LVKLQKAGLQFEGYENLPSDFN---LEDLKNIR---IKTPLSQDNKPKSLLDENNY 745	
DB	1413 YKELDEKY-----SEFNKTKIEEINNINRENINNVIEIWEYKNNIIEFLRHNDQK 1461	
QY	746 DGEIKAPFGLPLF-----LPKELR-----RNSNS-----G 772	
DB	1462 DKAAYME-NIDTYKNNIEISQINPENYVETLINKSNMYSYVEKANDLFYKQINNIIN 1520	

QY	773 GSQNSNRPW---EQEIIISQFKDQNLSDQOLAQFSTKIWEKIIGDENEFDQ---NNRL- 824	
DB	1521 SNQLKNEAFTIDELQNTQKRNLLTKKQIQIYTNEL-----ENFBIKININILV 1573	
QY	825 --QYK-LLKOLQESWINKTRDNLNWTYVLGDLKVKPKNNLEAKFQISNLQELTAFYTS 881	
DB	1574 LTNYSKILQDISQ---NINHVSIIYTEQL-----HNLVYKLEE--EKEQMKLYHKS 1619	
QY	882 AALSNMNYQDSGAKSTII--FEEIAELDPKVEKVGADVYQLKPHYAIGFDNAGKNQ 940	
DB	1620 NVLHQINFNEDAFINLLINIEKINDITHKEK--TNIYIMDVNKS---KNNALQYPH 1674	
QY	941 EVIRSSRTIYLTSGSKSLRADTIDOLNOAVKNAPIGLQSPFLDTERPGVFOKLATSLA 1000	
DB	1675 NTLKGNKIEYLNKLNKSNQOITLOELKQVQENVE-KVQDIYNOYTIKY----- 1722	
QY	1001 VQHKQKEXTLPKGLN--NDGYTLIHDKLKPKVI POISSPEKDWFEGLKNGQSQSN--- 1055	
DB	1723 -----BEKIKKNYHIITDYENKINDILHNSFIQINME-----SSNNKKQTKQI 1766	
QY	1056 ---VNVSTFGSIIIESPYSTNPOEDAD-----LQDQDQ---DSROGNSLDNQA 1100	
DB	1767 IDIINDKTFEHIKTSKTKINMLKEQSQMKIDKTLNEQALKLFVDINSTNNLNDMLS 1826	
QY	1101 GLLKQKLAILLGNQFIQYQQNDKEIB--FEINVEKVEL----- 1139	
DB	1827 EINS-----TONNIHTYIOBANKSDFPKIICDQNVNDLLNKLGLDNLNYYHLKNLQ 1880	
QY	1140 -----SPRVEFKL--AKTLENGKTIIRVLSDETSLIVNTTIEKTPEMSAVEPDTKWV 1192	
DB	1881 EIRNNMLEKNFMDLSKKIDSEKKDLILKYNISN--INNSLDLKKY--VEEALFQKVK 1936	
QY	1193 EYDPRTPLAATKTVLK-----FKQIIPVDS-----GNISD 1225	
DB	1937 EKAENQENIEKIOEINTLSDFPKFPFFQTLNTDSSQHEKDINNVTYKNNIDEIYN 1996	
QY	1226 KWLASIPLVIHQMLRLSPVW---KTIHGLKTEQOQQOQQOQQOQQOPOK----- 1273	
DB	1997 VFIQYNNLIQYSSSIFSTLNYIOTKEIKESIK--EQNLQNEKEASVLKNIKINET 2055	
QY	1274 ----KAVRKEELETYNPRDEFNINPLPLTKAHLTSLN-----VNNDPNYKIEDLKYI 1323	
DB	2056 IKLFQIKNERQNDVHNIEDYLLQYVLYNKMENEMQLKKYKNDVHMDKNY-----V 2108	
QY	1324 KNEADHQ-----LAFSLRANNIKRLMWTITADYNNPPFYNNEDWRSIDKLYANKGN 1376	
DB	2109 ENNNGEKKELKKTETISSYVDKINNINNKI-----YIYKKE---DTYFNMMIK 2153	
QY	1377 VS-----SHQQAAGNQGSGLIQRLNKNIKPETPTPALIALKRNNTLSNY 1424	
DB	2154 VSEIINIIKKQOQNEQRIVINAEDSSLINK--DESIKE--INNQIIEI--KHNENISNI 2210	
QY	1425 SKIIMIKPKYLVERSIGVPMSTGLDGYTGSQTKDGTSSSSQKQGFQDFTQALGLKNT 1484	
DB	2211 FKDIQNIKKQ-----SQDIITNM---N 2229	
QY	1485 EYHGLKGLSIRPDPONE-----LAKIDASNKKGB--EKLK----- 1520	
DB	2230 DMHKSTILLVDIIOKKEEALNKQKILNRIDNLNKKENIIDKVIKCNCDYKDILIQNE 2289	
QY	1521 -SYDLPKNVLYNEVKKSPKIAKGTWNIHPDQKEYPNPKLPENY-----LNLVLNQPW 1573	
DB	2290 TEYQKLQNLNHTYEKKKSI-----DILIKNIKQKNQYKQKLEQMTIINQSI 2340	
QY	1574 KYTLNSSDFITNLFVEPEGSRGSGTKLQVIQ-----KQVNNYADWGSAY-LTFWY 1626	
DB	2341 EQHVFINADILQN-----BKILEEIIKNDILDQELQIMTYHNSIDELYKLGIOG 2389	
QY	1627 DKNITNQNVITANADVFIDKVELEDNTKL 1659	
DB	2390 DNHLITTSVWVKNNTKIMIHKKQKEDIQKI 2422	

1271 POKKAVR-----KEBELTY-----NPKDEFNINPLTKAHLTIS 1306
1708 STKESIEPIIOMISFNSDLEDIASNALETYKTSMTNTYNEPMEYDILIKYSETVS 1767
1307 ---NLVNDPNVKIED---LAKVINEAGDHLAFSLRANNIKRL---MNTPIITFADYN 1355
1768 KEPTTYEDIKTRIEAQNELLNKNENNAKYLLODIDKANEFDRILVIYFMNKLNTLND-- 1825
1356 PFYFYNDWRSIDKYLNNKGNVSSHOOQAAGNOGSLIORLKN-----NIKPTF--- 1406
1826 ---KNSVIKIDEGLD---NISKYIDNVKSTDEDSLOKVLNQTKYELNIIKTYTSY 1879
1407 ---TPALIALKDRNNTNLSY-----SDKIMIKP---KY 1435
1880 KYDAONIPRNVKLSNLSNIQITNSGIDLSKNINAAISYLLNSPQNDPLIFVSPSLNKL 1939
1436 LVERSIGVWSTGLDGY-----IGSEOTKDGTSOSSQKGFDDFIQA 1478
1940 KYTKINNSYNTLLDIFKKSQDLHKKBOQTLDFGNRRLYETVQTTNELKTLND----- 1995
1479 LGLKNTYHGKLSIRIFDPGNEL-----AKIDASNKKGE 1515
1996 --LKNKENILNGVKL-VLDKSNIEIRQLSFGSONYDILEASKYNIQIKIDNYEKEQ 2052
1516 -----EKLKSYDLFRNYLNEYE-----KKSPKIAKWT-----NIHP 1548
2053 LRNFDDITWEEOLENIIKSIEOLENDYNSSENNHLOSKKLNBELTKTFNTEIKNIEN 2112
1549 DQKEYENPKLPEN-----YLVNLVNPQWVTLNSSFITN----- 1586
2113 KIEKNLISKLEDRKOCCLFTNTLEIEALKKNSNNYSEFTISATNFKSEFLKIYDTS 2172
1587 -----LFFV-----EPESDRGSG-TKL----- 1602
2173 NSLNDINNALQTKYNLINESKDVTRIFVDVNDNDNLDIKTEKATQAVSNLTKLFSIDLQ 2232
1603 -----KQVIOK-----QVNNYADWGSAYLTF 1624
2233 NADVNELYNNKLOMYFNSEAQSIDHIKQLYKILAFKLSNIDQINTKYSIDSLF----- 2289
1625 WYDKNIITQPNVITANI-----ADVFIDKVE--LEDNTKLIAPNIT---QWNPNI 1674
2290 ---DNIIOQKNKLTENLNLKETEYISDQKNFLHIRETINPOLNLTKEIYNNIIAR 2346
1675 KEFYKPTVPFGWENENSMNSQAQPTWEKIREGFALQALKSFDQKTRTPVLTTNAP 1734
2347 ENKIHB-IENINNEXDNITLYMDAITNLGKI-----QNLIDYVTI--- 2387
1735 LPLWKYGLFGONGPFPKTDQWRLVFNQD--DNQIAALRVQEODR-----PEKSSDQK 1787
2388 -----HENDHNITEQD---IQNDESNDKITEQDQIQNDKSDONITEQDQIDNDS 2434
1788 QKWIKKVVPIPEBMENSGNIRFVGWQIQGPNLWLPVINSVVIYDFYRGTSNDVANL 1847
2435 SDVSKIK-----NKLKTIOSLOQIQ-----QKI 2458
1848 NVAPWQVKTIATFTNNAFNPNFKBFNISK 1875
2459 N-----ETKTOFYDNNISINDIIT--TISK 2480
RESULT 23
Q6R5H8
ID Q6R5H8 PRELIMINARY; PRT; 1093 AA.
AC Q6R5H8; 2004 (TrEMBLrel. 27, Created)
DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE Adhesin.
OS Mycoplasma hyopneumoniae.
OC Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.
OX NCBI_TaxID=2099;

111 SEQUENCE FROM N.A.
RP Boisvert A., Cheikh Saad Bouh K., Khayar A., Shareck F., Dea S.;
RL Submitted (DEC-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY512904; AAS00639.1; -
DR InterPro; IPR000719; Prot kinase.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; UNKNOWN 1.
SQ SEQUENCE 1093 AA; 123337 MW; 3F883FB0B2E16AF0 CRC64;
Query Match 4.0%; Score 389; DB 2; Length 1093;
Best Local Similarity 20.6%; Pred. No. 8.3e-08;
Matches 262; Conservative 210; Mismatches 476; Indels 322; Gaps 59;
QY 1 MKNKSSTLLATAAALIGSTVGTGVLGASKVKYGVNPTQGVISQLGLIDSVAFKPSIA 60
DB 1 MSKSKSTFKIGTAGIVGLGVFGLTVGLSSLAKYRSFPRKIANDFAKVSTLAFSPYAF 60
QY 61 NFTSDYQVKKALNGKTFDPKSSSEFTDFVSKFDFLTNNGRTVLEI-----PKYQV 112
DB 61 ETDSDYKIVKRWLVDNNSN-NIRNKE--KVIDSFFTKNGDQLEKINFQDPVTKAKITF 117
QY 113 VISEFSPDDKXERFLGHLKEKLEDGNIAGSATFYIYLLPLDMPKAA---LGQYSYIVD 169
DB 118 EILEIIPDDVNVQNFVKFQALQKLHNGDIKSD---IYEQTVAFAKQSNLLVAEFNSLK 174
QY 170 KNFNLL-----IHLPLSNFS---AQSTKPLALTRESSDFIAKLQNFNNOELWVLEKFF 220
DB 175 KITEKLNQOENLSTKTIINFADKTSQKDPSTLRAIDFYDLNTARNAEDLDIKLANVF 234
QY 221 DLEALKANI-RIQTADFSEFEGNLVDPFVYGFIRNPQNKQWASDLNQDQKTVRLYLRT 279
DB 235 PV--LKNLINLNNAPENKLPNNLGNIFEFGAKD--SSTNQYVSIQNG---IPSLFLKAD 288
QY 280 FSPQAKTILKDYKQDET--FLSSIDLKASNTSLFANENDLKQDLVDLDDVSDYFGQ 337
DB 289 LSQSAEILAS---PDEVQPVNIIRLMKQNSSYFLNFDVFNLL----- 331
QY 338 SETITNSQVQVPVPSASRLKDRVKFKDQKQPRTEKFSLEY-----DALSPY 386
DB 332 -----TLKN---MOKEDLNKQNLISAYEFLADIKSGFPFGKRSR 370
QY 387 --SQQLVSKPNSIKDLVATLARNLAFSLGKYNPLFDDL---ASHLDYFVLVSKAKIKQ 442
DB 371 TKABEISNLLNKKENIYD-----FCKYNGKFNDRLNSPYLEYSLDAASASLDK 417
QY 443 SSITKKLFTLPIKISLSKSIIGDQ-EPNIK-TLFEKEVTFKLDNPRDVEIEKAFGLLYP 500
DB 418 KD---KSVILIPYRLKIDKFPADDLYPDTQDNILVKEGILKLTGPK-----KGPKIDLP 469
QY 501 GYNESLEQAKAQRASFEKEKSKGKBPSCQKEENSKAINNOEGLEEDDNITERLPENS 560
DB 470 NINQOI-----FKTEY---LPFEKKEEQAKL-----DYGNI- 500
QY 561 PIQYQOENAGLGPDPKPYMIKDVQNYRYLLAKSQIQELIKAKDYTKLAKLNSRHITYMI 620
DB 501 -----NPYNTQ-----LAKVEALFKGNQKEIYQALDGNVAYEF 536
QY 621 SURLEKQLPDVNPRI--PSSRDIEK-----AKFVLKTEKGNQWQIYSS--ASP 665
DB 537 G-AFKSVLNSWTGKIQHPEKADIQRFTRHLHQVKIGSNVLPNQPTAKE-QVSSLSKSN 594
QY 666 VFQNKWSLFGYVRYLLGLDPKQTIHELKVLQKQAGL-----QF-EGYNLPSDFNLEDLK 719
DB 595 FFGNGHQVASYFDLLTKDKLTVLETLYDLAKKNGLETNQOAFPGAFQYTKDIPAEADK 654
QY 720 NIRIKTFLPSQKDNF-----KLSLLDFNNYYDGE-IPKAFEGFLPLFLPKBLRNNSSNGS 774
DB 655 ---LKFLEWKKDPYQYQINEIHQLSFNILARNVDVIKSDGFGYVLLLPQSVKTELEGN-- 709
QY 775 QNSNSPWEQEIISQFKQKQNLQDQALQFSTKIWEKIIGDENEFDQ--NNRLQYKLLKDLQ 833
DB 710 -----EAQIFEALKKYSLEIEN---SAFKTTILDKNLLEGTDFKTFGDFLKAFFLAAQ 759


```
QY 1006 KECTLPKLN--NDGYTLIHDKKLPVIOPISSPEKDWFGKLNQOSQ--VN 1057
DB 1712 -EEBKKYVHIITDYENKINILHNSFFIKQINME-----SSNNKQTKQIIDIIN 1760
QY 1058 VSTFGSITESPYFTSNFOEDAD-----LDQGDQ--DSRQGNNSLNDQAGLLKQ 1105
DB 1761 DKTFEEHKTSTKINMLKEOSQMKHIDKTLNEQALKLPVDINSNNLNDMLSEINS- 1819
QY 1106 KLAILLGNQFIQYQONDKETB-FRIINVEKVEL----- 1139
DB 1820 -----IQNNIITYIQEANKSFDKFIICDQNVNDLLNKLGLDLYNMHLKNQLNERNM 1874
QY 1140 SFRVEFKL--ARTLEDNGKTIIRVLS-----DETMSLIVNTTIEKTEPMSAVPEVF 1187
DB 1875 NLEKNFMLDKSKIDEEBKLDILKVNISNINNSLDLKYVEEALFOKVKKEAIOKEN 1934
QY 1188 DTQWVEQVDRPTPLAAKTKFVLKFKQIIPVDS-----GNISDKWLAS 1230
DB 1935 IEKIQEINTLSDVPKPPFFI---QLNTDSSQHEKDINNNAVETYKNNIDEIYNVFQS 1990
QY 1231 IPLVTHQMLRLSPV-----KTIRELGLKTEQQQQQQQQQQQQPOK-----K 1274
DB 1991 YNLQKYSSEFSSTLNIQYKBEKSIK-EQNQLNQNEKEASVLLKNIKINETIKLKF 2049
QY 1275 AVRKEELETYNPKDEFNINPLTKAHLRTLNLVNDPNYKIE---DLKVINKNEAGDHQ 1331
DB 2050 QIKNERQNDVNIKEDYNLL---QQYLNVMKNEQEQLKKYKIDVHMDKNVYNNNGEKE 2105
QY 1332 -----LAFSLRANNIKRLMNTPTTFADYNPFYFYNEDWRSIDKYLNNKNGVS----- 1378
DB 2106 KLLKETISSYDVKINNNKLL-----YIYKKE---DITYFNMMIKVSEILNII 2150
QY 1379 -----SHQQAAGNQGSGIOLRNKNIKPTTFPALIALKDRNNTLSNYSKIIIMIK 1432
DB 2151 IKKQONQORIVINAYDSSLNK-DEEIKKE-INNQIIEI-LNKNENISNIFDIQNIK 2207
QY 1433 PKYLVERSIGVPMSTGLDYGSEQTKDGTSSSQKQFGDQFTIQAALGLKNTYHGKLG 1492
DB 2208 KQ-----SQDIITNMDCMCKSTILLVDIIQ---KKEEALNK--- 2240
QY 1493 SIRIPDPGNEAKIDASNKKE--EKLLK-----SYDLFKYLNVEYKK 1535
DB 2241 -----QKNILRNIDNILNKKENIDKVIKNCDDYKDLIQNETEYQKLNINHTYEK 2294
QY 1536 SPKIAKGTWNIHPDQKEYPNQKLPENY-----LNLVLNQPKVTLNNSDFITNLFV 1589
DB 2295 KKS-----DILKIKNIKQKNIOYKKNLQOMNTIINQSIQHVFINADILQN--- 2342
QY 1590 EPGSDRSGTKLKQVIO-----KQVNNNADWGSAY-LTFWYDKNIITNQPNVITANI 1642
DB 2343 -----EKIKLEBIIKNLDLBEQIMTYHNSIDELYKLGICQDNHLITTSVNVNKN 2394
QY 1643 ADVFIKDVKELEDNTKL 1659
DB 2395 TKIMIHIKKQKEDIQKI 2411

RESULT 26
Q7RJ53 PRELIMINARY; PRT; 2261 AA.
AC Q7RJ53
ID Q7RJ53;
DT 01-MAR-2004 (TReMBLrel. 26, Created)
DT 01-MAR-2004 (TReMBLrel. 26, Last sequence update)
DE 01-MAR-2004 (TReMBLrel. 26, Last annotation update)
DE 235 kDa rhotry protein (Fragment).
GN Name=pf03184;
OS Plasmodium yoelii yoelii.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=73239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=17XNL;
```

```
RR PubMed=12369865; DOI=10.1038/nature01099;
RA Carlton J.M., Angiuoli S.V., Suh B.B., Kooij T.W., Pertea M.,
RA Silva J.C., Ermolaeva M.D., Allen J.E., Selengut J.D., Koo H.L.,
RA Peterson J.D., Pop M., Koeack D.S., Shumway M.F., Bidwell S.L.,
RA Shalom S.J., van Aken S.B., Riedmuller S.B., Feldblyum T.V.,
RA Cho J.K., Quackenbush J., Sedegan M., Shoabi A., Cummings L.M.,
RA Florens L., Yates F.R. III, Raine J.D., Sinden R.E., Harris M.A.,
RA Cunningham D.A., Preiser P.R., Bergman L.W., Vaidya A.B.,
RA van Lin L.H., Janse C.J., Waters A.P., Smith H.O., White O.R.,
RA Salzberg S.L., Venter J.C., Fraser C.M., Hoffman S.L., Gardner M.J.,
RA Carucci D.J.;
RT "Genome sequence and comparative analysis of the model rodent malaria
RT parasite Plasmodium yoelii yoelii."
RL Nature 419:512-519(2002).
CC -|- CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
DR EMBL; AAEU01000904; EAA22730.1; -.
DR InterPro; IPR003900; KID repeat.
DR InterPro; IPR006499; ReticulocyteBP.
DR Pfam; PF02524; KID; 3.
DR TIGRFAMs; TIGR01612; 235kDa-fam; 1.
DR NON_TER 1
FT NON_TER 2261 2261
SQ SEQUENCE 2261 AA; 264846 MW; B7BE8C72B13BC091 CRC64;

Query Match 4.0%; Score 385.5; DB 2; Length 2261;
Best Local Similarity 19.7%; Pred. No. 3e-07;
Matches 418; Conservative 334; Mismatches 724; Indels 641; Gaps 106;

QY 82 KSSEFTDFVSKDFLNNNGRTV-----LEIPKQYQVISEFSPEDDKERFLGFLHKE 134
DB 315 KNSLKNLVTKLEFIKKQNEIDIQNHKHIETCKBEVSVNKT-----KT 360
QY 135 KLEDGNIQATKFIYLLPLDMPKALGOVSYLVKDNFNLLIHPISNESAQSTKPLALT 194
DB 361 IINTNCDQHFEI-----KKIGE--SYIIK--YHHTVLEYL-----ESIR--ITYK 402
QY 195 RSSDFIAKL-----NQFNQDELWVYLEKFFLEALKANIRLQATDFSEKGNLVDPFV 248
DB 403 HSLYFLKLGLDALINKVDSDGNI-IEKDISDFNLLKQSTFKSLEAEFLK---IFENK 458
QY 249 YSIRPNQ-----QKEWAS-----DLNQDQ---KTVRLYLRTEFSPQAKTIL 288
DB 459 WDFYKNTQNLTKDPMKNTISLILPLNQFKGLNESMIKLNKSGILRKTYISNQIKNL 518
QY 289 KDVKYKD--ETFLSSIDL-KASNGTSL-----FANENDLKQDLQDVLDDV--SDYFGQSE 339
DB 519 NVSTYBGRGFTSSLELAKSWEKTKLETTITELTKSNEETVRLEKEIRELFKYLDEAE 578
QY 340 -----TITSNQKVPVPASERSLKDRVKFKDQKQPR----- 371
DB 579 RYKLEGLKLELNKKIDIIAKIEVKNVTELKKEIEKNNAVIDELANQSPVKVTGYIENK 638
QY 372 -----IEKF--SLYEYDALSFYSQLOELVSKNSIKDLVNATLARNLRLSLGKYNLF 423
DB 639 NTIYTIKSYFDQIYEGDIDTFYNELSSIV-KEDPIDDIEDTKLENLR---SKIDNVYD 694
QY 424 DLASHLDYYFLVSKAKIKQSSITKLFELPIKLSKSSILGQEPNIKTLFE----- 476
DB 695 KI-----QKMEIE--TVKSHLNNIETNNKLPN--TILEIKKIYD 730
QY 477 ---KEVTFKLDNFRDVEIEKAFGLLYPGVNEELEQAKAQAQASFEKSKKGLKEFSQK 533
DB 731 EISKELNKMLEDKF-----NKEKELSNKIS---DYDKGRELSEYKSKM 771
QY 534 EENSKAINNOEGLEEDNITERLPENSPIOYQENAGLGPSPKPYMIKDVQNRVYLAK 593
DB 772 LEIRNNHNSQTNV---DNTKE---EEAKQYDKSNEHMTTIPITNEDEISKI----- 816
QY 594 SQIOELIKAKDYTKLAKLNSNRHTY-NISLRLEQLFDVNPRIPIPSSRDIEKAKFLVDKTE 652
DB 817 --ISEVKTMD-----EILSKVNTYIDFNKKYKETVNSEHSQFTELTDKIKAE--VSDKEL 868
```


Db 507 VIDFMENTQDVEPTRRIIGPEVEWRIKALL-----YSEVAGVKKSYESVKNWRKS 558
Qy 103 VLEIPKXYQVVISPEPDDKXERPLGLHKE-----KLEDGNIQASATKFIYLLPLDM 156
Db 559 ILEINKKLE-----BENEKVIKLETQIKDLFNQYLIKIDEMI-----YINKLKLEL 604
Qy 157 PKAALGOYSYIVDKN-----FNNLIHPLSNFSAOSIKPLALTRSSDFI-- 200
Db 605 KE-----KIKNISDRKVEYIKAVOLKIIENNAYIDELAKTSPYQV-PEHL-KSTDIIYS 658
Qy 201 ---AKLQFNQDELWVYLE-----KPFOLEALKANI--RLQTADESP 238
Db 659 TIKSLSQIYEDDIDKLYNELSSIVQENDIDNVEDKTKLDLQSKIDNVSQIKQ----- 712
Qy 239 EKNGLVDFPVYFIRNPQNKQEWASDLNODQKTVRLYLRTFSPQAKTILADYKQDET 298
Db 713 ---NMETATVSSHLTNIETNKNKUSDITVE---IKKYIYGEISKDLNKTLEDFKNKEK-- 764
Qy 299 LSSIDLKASNGTSLPANEND-----LKDQLDVIDLDVSD-----YFGQSE 339
Db 765 -----ELSNKINDYAKENDQLNVYKSISEIRNHYNQDISIDNTKEBEAKQNY--DQSK 816
Qy 340 TITSNQVKVPAPASERSLKDRVKFKDQOKPRIEFKFSLEYEDAL-----SFYSQLOELVSK 395
Db 817 EYTKTISIKEDETSXTI--NEVQNMKDFLGSVDKYINFENCKENVDSEHTQFTLTNK 874
Qy 396 P-----NSIKDVLNATLARNLRFSLGKYNFLEDDLASHLDVYFLV--S 436
Db 875 IKAESVDEKLSIYENKFNDSKSLNET--NNGIEKEYQINIL-----KKVDEYIKVCEST 928
Qy 437 KAKIKOSSITRKLFIELPIKISLKSILGDOQ-EPNIKTLPK-----EVTFKLDNFRDV-- 489
Db 929 KESIK-----NFHNQITILKEKLQNQIKTIKETNSIEKSYK-DKPENTLIN 973
Qy 490 ---EIEKAFGLL-----YGVNBELEQAKQASFEKSKGLKFPQOQKEBSKAINN 542
Db 974 KINELDKAFKQASLDYESSNNELMQYFNNLKANLGNKNENMLYHQF-----DEKEKAVNN 1029
Qy 543 QEGLEEDDNIETRLPENSPIQYQBNAGLSPDKPYMKDVQNO-----RYVLAK 593
Db 1030 I--IOKIEDINKNVIP-----NIBIAIYTSIYNISSIEIENIGKIESLNTNVLEK 1077
Qy 594 -----SOIQELIKAKDYTKLAKLLSNRHTYINSLRKEQLFDVNPRI----- 635
Db 1078 VNTNVTNFKIKEKLYNFSDFGKEGNIKYTNEVN-KIKODIKVNAQOIHINELEEI 1136
Qy 636 ---PSRDIEKAKFVLDKTEKKNYQIYSSASPVQNKWSLFGYRYLLGLDPKQTIHEL 693
Db 1137 KKKSENYVDEKQAQINKLEKVAADTAISNDNVEGIEK-----QQNI VTKIDKKXNIYEBI 1191
Qy 694 -----KLQKAGLQPEGYENLPSPDNLEDL-----KNIRIKTLPFSQ 730
Db 1192 NKLLSEISKIEKQOTSLEKVGINLSYQNLGTLF-LEQIDBEKKSEHTIKAMEAYMED 1250
Qy 731 KDNFKLSLLDFNNYVDGEIKAPFGLPLFLPKELR-RNSSNSGG-----SONSNSPWEQ 783
Db 1251 LDNIKKKSQEIEN-----EMGIEMDINKEMEVLNISHDDOKKYHTISKHNN----- 1296
Qy 784 ELISQFKQNLNSQDQLAQFSTKIWEKLIIGDENPQNNRLOYKLLKDLQESWINKTRDN 843
Db 1297 ENISDIRKSL-----KIIQDFRESNDINDIKKELQKRVSSQKHNSIN 1341
Qy 844 LYWTYLGDKLVKPKKNLEAKFROIISNLOELLTAPYTSAAALSNWNNYQDSKASTIIFE 903
Db 1342 QYLSKIANIYNILKLNKIK--KIINKVKE-----YTNEIEKNKQV-----IND 1382
Qy 904 EIAELDPKVEKVGADVYQLKPHYAIGFDNAGKPNQBNVIRSSRTIYLTSGSKLEAD 963
Db 1383 ELSN-----SEKLIKIKDIDSLKEC-KSKIES- 1409
Qy 964 TIDQLNQAVKNAPLGLQSFYLDTERFGVFOKLATSLAVQHKQKETLPKLLNNQGYTLIH 1023

Db 1410 TLD-----DKDIDGCIQNIKV-----LKTHILSEETNINTYFNQA 1444
Qy 1024 DKLKPKVI---POISSSPKDFEGKLNQNG---QSONVNVSTFGSIIESPSTNPFQ 1076
Db 1445 DEYNKVNLLNPNLEIMADNKSQVILKIKONGTNDHDYNN-----ELKEHKDKSNGYKT 1499
Qy 1077 DADLDQDQDSDSQNGNSLNDQBAGLQKLAILLGNQFOIYYOQNDKBEIEFEI----- 1131
Db 1500 EADKNKKAIEKNKELFQYQEVTVLLNKYVAVELKNKFDK--TKNDSE---QIIKEIKD 1554
Qy 1132 -----NVEKVEL---SPRVEFKLAKTLEDNGKTRVLSDETMSLIVNTTIEKT 1177
Db 1555 AHNKFILEAGKSQOKMEIKNEQIRIEDEKAK---NDKSNKAIIDIQLSV----- 1601
Qy 1178 PMSAVEVEPDTKWEQYDPRTPLAAKTKFVKDQIPVDGSGNISDKWLASIPLVHIQ 1237
Db 1602 -----EPETFKLISDIRT---KSNDCUKETESI-----EKKISTISLIDTOE 1641
Qy 1238 QMLRLS--PVVKTIREL--GLKTBQOQOQOQOQOQOQOQKQKAVRKEBELETYNPKDEFNI- 1293
Db 1642 TKLKGENDLNTLQEFLESKQKQKNIEDKKELDEVNSKIEKIKIDVNOHKKNYEIGIV 1701
Qy 1294 --LNPLTKAH-----RLTSLNVNN-----DPNYKI-----EDLKVIKNEA 1327
Db 1702 EKINEIAKANNKIEBTKELIKPTIERLISSNTNDLEGIDPNKGEYNTMGGIITNEF 1761
Qy 1328 GDHQLAFSLRANNIKRLMNTPIITPAD-----YNPFYFYNEDWRSIDKYL-----NNK 1374
Db 1762 ME---LYNLGAVLKTYSKESIYNEIKNTRINAQNELLNKIESEKSKSYLDDIEANEF 1818
Qy 1375 GNVSQHQOQAGG-----NOGSLIQRLNKNVIRPETTPPALIAIKORNNTNLS 1422
Db 1819 DRIVTHLKKLNDVNVKFTNEYSKVNEGPNISKSINNKNSTDENSLNLTNLTREMYA 1878
Qy 1423 N-----YS-----DKIIMIKPKYLVERSIGVPMSTGLD-----GVIGSEQTKDG 1461
Db 1879 NIVSKYYSYKYEAEKIFINISKLANSNIQIQNSSGVDDLKNINIAILSYLDSFKEDTL 1938
Qy 1462 TSSSSQOQKGF-----NTEBKKIEBQIIEKNLIDINKLIEITRQCLLFTVTTL 1472
Db 1939 TFTPSPQKSETYKIRDSYNTLLDIFKSOYLHKKEQQTLLNLI FENRRLYEKVQATNEL 1998
Qy 1473 QDFTQALGLKNTEYHKGGLSIRIFDPGNEELAKIKDASNNKGBEKLKS--YDLFKQVYN 1530
Db 1999 KDTLSDLKYKKEKILNEVKL---LLHKSNEKLNKL--SCNSQNYDITILESSKYDKIKEKSN 2053
Qy 1531 EYEKSKPKIAGKWTNIHPDOKEYPN---PNQKLPENYLNVLNQPWKVLYNNSD----- 1582
Db 2054 NYEKEKENLGINF-DVAMEEQFNNDIKDIEKLENNYKHSEKN-----YNFSEENNI 2106
Qy 1583 -----FITNLFVEPEGSGRSGTKLK---QVIQK-QVNNNYADWGSAYLTFWYD-- 1627
Db 2107 LQSKKKLKELTNAF-----NTEBKKIEBQIIEKNLIDINKLIEITRQCLLFTVTTL 2156
Qy 1628 ---KNITITQPNVITANIADVFKD--VKELEONTKLIAPNIQOWBNISSSSEKFEKPYKT 1682
Db 2157 VETLSKVTHSEFITS--AAKFSKDFKYIEDISNLSNDIDITLQINYNLTNQTNGKVP 2214
Qy 1683 VFFGNWENENSNMSOAO-TPTWEKIREGFALOALKSSFD--QKTRFTVLTNAPL----- 1735
Db 2215 ILADATKOHNNLKEKEKATKIINNLTETIDSNNDIADTLHNNKIQIIFYNSLHLSKI 2274
Qy 1736 -----PLWKYGPLGFONGPNFKTQDMRLVFQNDNOIAALRVQEQDRPKSEKSDK 1785
Db 2275 ESIKQLYKKNVIFKLLNISHINGKYFD-----ISKEFDN---TLOMQKHKLJENLANDLK 2325
Qy 1786 DKQWIKFEKVVIFPEEMENSGNIRFVGVM-QIOGNTLWLPVINSVYDFVRGTGSDNDV 1844
Db 2326 QIOQYISDK-----KNKFLHALNETTNPNTLKEIYHDIIV-NYERQJLDEIKNI 2373
Qy 1845 ANL---NVAPW-----QVKTI-----AFTNNAFNVPFKE 1870
Db 2374 SNKENENITSYIDTITKLTGKVQNLNLFVITYENN--NNIIKQ 2414

Db 2225 DISIFNKY-----DDYIKVDYMSNNIDVV-----NKGNSLLSEHVINATNIE 2268
Qy 1404 EFTFPALIALKORNTNLSYGDKIMI-----KPKYLVERSIGVPMWSTGLDGYIGSBOT 1458
Db 2269 NIWTCIVQINKDTEMNFTBETKDNLLKLYENFKKEKNIINNNYKIVHFNKJKEIENSLET 2328
Qy 1459 KDGTSSSQOQFDDPFIQALGKNTYHGKGLSIRIFDPGNELA-----1504
Db 2329 YNSISTNFKISAQN-----IGLKNFEFN-----NITKINDKVKELVHVVDSTUTLESIQTF 2381
Qy 1505-----KIKDASNKKGKELKSY-----DLFKNYLNEYKKSPP 1538
Db 2382 NKLYGDMGNIQDFYKYEINN--GELKKVLYIENITNLLGRINTFIKELNKYQDENVV 2439
Qy 1539 IAKGWNIIHPDQKEY-----PNPNQKLPENYMLVLNQPWKT--LYNSS-----DFITNLFVE 1590
Db 2440 IDK-YTEINKNNSYIIKLEKANNLKQFNLLQNIKENETELYNINSIKODIMNMWKS 2498
Qy 1591 PEGSDR--GSGTKLKQVI-----QKQVNNYADMGSAYLTFWYDQNIIT 1632
Db 2499 VNNIKQKFRNLPLKEKLFQMEKMLDINNIMNETKRISNTDAYTNTIQTDIENNKNKEN 2558
Qy 1633 NOPNVTANIADVFIDKVKLELNDNTKLIAPNITQWMPNISGSKKFKYKPTVFFGNWENEN 1692
Db 2559 NNNIKT--IDKLIDDIK--IHNEKIQAEILI-----IDDAKRYNEIT-----DNIN 2602
Qy 1693 SSMNSQAQPTTWBKIREGFALQAKSSFDQKTRFTVLTNAPLPLWKYGLFGFQNGPNFK 1752
Db 2603 KAPTEITETNNND--NNGVIKSAKNIVDEAT---YLNNELOKFLKLKLNELLSHNNNDIK 2656
Qy 1753 -TQDWLTVQNDNDQTAALRVQODRPEKSSSEDKQK 1789
Db 2657 DLDDEKLIILKBELEQKLAQBEERKERERIEKEKOE 2694

RESULT 32
Q26223 PRELIMINARY; . PRT; 2269 AA.
AC Q26223;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Rhostry protein.
OS Plasmodium yoelii.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5861;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=YM.
RX MEDLINE=95021522; PubMed=7935623; DOI=10.1016/0166-6851(94)90125-2;
RA Keen J., Sinha K., Brown K., Holder A.;
RT "A gene coding for a high molecular mass rhostry protein of Plasmodium
RT yoelii".
RL Mol. Biochem. Parasitol. 65:171-177 (1994).
DR EMBL; L27838; AAA21304.1; -.
DR PIR; T28677; T28677.
DR HSSP; P02652; 116L.
DR InterPro; IPR006499; ReticulocyteBP.
DR TrIRPAMS; TIGR01612; 235kDa-fam; 1.
SQ SEQUENCE 2269 AA; 265158 MW; F3D8CB103FB9A6E1 CRC64;

Query Match 3.9%; Score 377; DB 2; Length 2269;
Best Local Similarity 18.5%; Pred. No. 6.7e-07;
Matches 418; Conservative 363; Mismatches 762; Indels 712; Gaps 106;
Qy 19 STVPGTVGLASKVKYRG-----VNPTQGVISQLGLDLSVAPKPSIANFTSDYQSVKAL 73
Db 165 NTIYNTIKSYDQI-YEGDIDTTFNELSSIVKE-DPIDDIEDTKLENLSKIDNVYDKI 222
Qy 74 LNKGTDPKSESEFTDFVSKFDFLTNNGR---TVLEIPKQYQVVISSEPS- 120
Db 223 -----QKWEIETVKSLLNNITNNKLPNTILEI-KKY--IYDEISKELNKLWEDFKN 271

Qy 121 -----DDKERFRLGPHLKEKLEDGHIASQATKFIYLLPLDPMKAALGOYSYIVD 169
Db 272 KEKELSNKISDYDKKREQLSEYKSKMLEIRHNYSQTN-----VDNTKEBEAKQNY--D 323
Qy 170 KNFNMLIIHPLSNFSAQSIKPLALTRSSDFIAKLN---OFNNQDELWYLYE--KPFDL-E 223
Db 324 KSNEHMTIPTNEIBEISKIIGEVKTMKDBILSKVNTYIDFNKKYKETYVNSEHSQFTELTD 383
Qy 224 ALKANI---RLQADPFSEKGNLVDPFVYSTRPNQ--QKEWASDLADODOKTVLYLRL- 277
Db 384 KIKAEVSKELKCKEQSFNDK-----SLINETKNSIEKEY-QNIN--TLKKVDEYIKV 434
Qy 278 -----TEFSQAKTILKD-----YKTKDETFLSSIDLKASNGTS 311
Db 435 CKSTKESITKSSK-OTILKQMLNQNIKTVKETNSIDKSYIEKFPQILLTGQTKLENKFT 493
Qy 312 LP-----ANENDLKQQLDVLDDVSDVFGGSEITITSNSQVQVPASERSLKDVRKFKK 365
Db 494 EPLSNHNEANNELIKYFS---DLKANLGINENMLYNQFTE---KEKTFND-IKEKN 544
Qy 366 DOOKPRIEKSLEYDALSFYSQLQELYSKNSIKDLAVNATLARNLRSLSGKYNFLDDEL 425
Db 545 IHINEBISKIETIKH--ASIYNISEETREBIGINIESLNTKVFKEKGVNTVNLNKKI KEKL 602
Qy 426 ASHLDYVYPLVSKAKIKQSSITKCL---FIELPIKISLSSIIIGDOEPNIKTLFE--KEVT 480
Db 603 -KHVDFSGEGEKNIKYTDKIKINDDIMASQQIDQIHGLDDIQKSESIVSEMKQI 661
Qy 481 FKLDNPRDVEI--EKAQGL-----LYPGVNEBELQOARKAQAASFEKES 522
Db 662 NKLEKVSNTESINDNVEGIKKKQIIVTKIDKKNIYBEINKLLSEISKIEXDNTSLEKV 721
Qy 523 KK-----GLKEFSQOKEKSAKNNOGLEED-DNITERLPE-----NSPI 562
Db 722 KDINLSYQGNLNFLEQIDEEKKAENTIKSMEAYIDDLNKKKQSEIETEMDKMDI 781
Qy 563 QYQENAGLGSADPKPYMIKDVQNR-----YLLAKSQIQLIKAKDY---TKLAKLS 613
Db 782 NKEMALKISHDDDKCHDKSKNHNENISDIYDKSKIIQDFSRESDDINDIKLKQKWS 841
Qy 614 NRHTYNISRLKELQFDVNPRISSRDIEKAKFVLDKTEKNKWKYQIYSSASPVQKNSL 673
Db 842 ESONHNSDI---NOCLNEVANIYNILKLNKIKKIIDKVKES-----YTSIEKN- 886
Qy 674 FGYYRYLLGLDPKQTHLKVGLQKAGLQFAGVENLPS-----DFNLED-----LKNIR 722
Db 887 -----KQINDELNNSEKVIKIBGDLSLKECRSKINSLTDDDKDIDECIKNIN 934
Qy 723 I-KTPLFSOKNF-----KLSLDDNN-----YVDGEIKA-----PEPGL 756
Db 935 VLKKNILNEETNITNHFKNABEYKIVLSNPANNIEMADNKSQYILEIKKNGTNDHDYNI 994
Qy 757 PLFLPKELARNSSNGG-----SONSNS--PMQETISQFDQNLSNQDQLAOFSTKWEK 810
Db 995 -----KELSKHDKSGNGYKTEADQNKKAIQKNKELPEQYKEE-----VTLLNK 1038
Qy 811 IIGDE--NEFDQNNRLQYKLLKDLQESW-----INKT 840
Db 1039 YVAVELKKNKFDTKNDKSKQIIEIKDAHNYCTLESKSEKKNKNEIKHIEDEVAND 1098
Qy 841 RNLVWYTLGDKLVKVP-----KNNLEAKPROISNL----- 871
Db 1099 KSNKAITSI--KVSVEPFTKIIKINEITRSDCLKETNDLE---KQISNLSIDTQETK 1153
Qy 872 -----QELLTAFYTSAAALNNNNYQDSGARSTIIFERIAELDPKVKYKGVADYQKPH 926
Db 1154 LTENGQKLTLELLESLKQKNEDQK-----KELDEVNSKIK-NIENTVNOHKKN 1205
Qy 927 YAIGF-----DDNAGKN----- 939
Db 1206 YEIGVEKINEIAKTNNKQIESTKELIKPTIOHIISSFNANDLEGDSIDSGKYNTMG 1265
Qy 940 ---QEVIRSSSR-TIYLTSGK-----SKLEADTIDQLNQAVKNAP-LGIQSFYLDTERF 989

Db 1061 INKLADEISE-----IEKNNTSLKKVKDINVYQSGSLGKLFLE-----QIDEKKKAGHM 1110
Qy 468 EBNITKLPKVTFKLDNPR-----DVEIEKAFGL-----LYPGVNEELE----- 507
Db 1111 ---IKSM---EYMDLDNLIKKSQBIENEMKINMDIKMY--INKEWALNISHDDDKQVH 1163
Qy 508 --QAKAQAASPEKESKGLKEFS-----QOKEENSKAINV----- 542
Db 1164 NISKHBEKISDIHKNSLKIIIEFESTESNINDIKNELOKVSESRYNNNDINQIILSKIAS 1223
Qy 543 -----QEGLEEDNITELPENSPIYOQEN-----AGLCA 573
Db 1224 IYNILKLNKIKLIIDKVKETHEIEKNKNITVELNSKTIIELEDDSNLKDCKSKIGS 1283
Qy 574 SPDKPYM---IKDVQNRQRYLAKSQIOELIKARDYTKLAKLSNRHTYINISLRKEQLFD 630
Db 1284 TIDDDVVGSCIKKIDTKTYILNKE-----TNIAHFNAEKYENVLL----- 1327
Qy 631 VNPRISSRDIE-----KAKFVLDKTEKKNYQWYISSASPVFQNKWSLFGYRYLLGLDPK 686
Db 1328 -----NPNNIEMADNKAQYIL-KIKKN-----NGTNHHDY----- 1356
Qy 687 QTHLVLGQKAGLQFEGVENLPSEDFLELDKNIRIKTLPFSOKDNFKLSLLDFNNYD 746
Db 1357 -NITELKKGKOMS-----NGYKN-EADTNIQAFK---NKELEKYKQDVTVLDD--KYV- 1404
Qy 747 GEIKAPEFGLPLPKELRRNSNGSGSQNSPMBEQIISQFQON-----LSNQDLAQF 803
Db 1405 -----AJALKNKFDKNDK-----EQIIEKIDHTHTFFISOADKSEOK 1443
Qy 804 STKIWEKIDGENEFQNNR-----LQYKLLKDOLESWINKTRD-- 842
Db 1444 MKEIKNEKINIEDEVANDKNSKAIIDIQLSVQPFETKFLKIDKLTSDDCCLKTEKOTE 1503
Qy 843 ----NLYWTYLGDKLVKPNLEAKFROISLNOELLTAPYTSAAALSNNWNYQDSGAKS 898
Db 1504 KEISNLSI-----DTQETKLENGE-----KLDTEKLI-----ESLKNKKYIE--GKK- 1546
Qy 899 TIIFEIEAELDPKVEKGVADVYQLKPHYAIGF----- 931
Db 1547 ----KELDEVNSKI-EKIKIDVKNKKNVEIGIVEKINEIAKANKQIESTQKLIKPTIE 1601
Qy 932 -----DDNAGKNOEV-----IRSSRTI-YLKTSGKSKLEADTIDQ 967
Db 1602 KLTSSPKTNDLKGINTDENLKGNTGEMGVINEFMESYNRIADYLTKTVSKPEITYDDI-- 1659
Qy 968 LNOAVKNAPLGLQSFYLDTERFGVFOKLATSLAVOHKOKE-----KTLPKK 1013
Db 1660 ----XNTRINAQN-----ELSKIKIENKAKSYLNDIEANBFDRIVTHFKNK 1702
Qy 1014 LN--NDGYTLIHDKJLKKPV-----IPIQISSPEKDWFEKLNQNGSQSNVNVSTFGSII 1065
Db 1703 LNTVNDNFKNEYKVNEGPDNISINNVKNSTDENLLNMLNQTQO-----IVANIV 1755
Qy 1066 ESPYFTNFQEDADLDQDQDQSDRQGNLSLDQEAQGLKQKLAILLGNQFIQYVQNDKE 1125
Db 1756 NQYYSYKYEAB-----NIFIN-----IPKLANTLNIQI-----KNSSG 1789
Qy 1126 IE-FEILNEKVSSELSFRVEFKL--AKTLEDNGKTIRVLSDEMTSLVNTTIEKTPMSA 1182
Db 1790 IDLFGKNINAILPHLSQKEDTLTFSPSPHNSIKTYTKISDSYNTLL--DILKRSQELQK 1847
Qy 1183 VPE-----VPDTRKVEQYDPRTPLAARKFVLKFKQIQIPVDGSGNISDRKLASIPVHQ 1237
Db 1848 KEQKTLNLIENRLL--YD-----KQVATNELKOTL--SNLKNKKEQILNEVKLLLHK 1896
Qy 1238 --OMLRLS-----PVVKTIRELGLKTEQOQQOQQOQQO-----QOQPKAVRKE 1279
Db 1897 SNEINLKLSCNFQYDALEPAKYDKIKEKSNYEQEKNKLGIDFDVTAMEKFPNNDIKVI 1956
Qy 1280 EELE-----TYPKDE-FNINPLTKAHLRTLSNLVNDPNPNKYIEDLKVKEAGD 1329
Db 1957 EELENNYKSEKNTYNFSENNILQSKQKLKELT--NEFNTEIK-RIEDKVIEKN--GL 2011

Qy 1330 HOLAFSLRAN-----NIKRLMTPITPAFYNPFF-----YYNEDWPSIDKYLNNKGN- 1376
Db 2012 INKLEIRKNCMLPHTHTLSETLKIKIT--DYSEFIASATKFSKFLYIDDISNSLND 2069
Qy 1377 -----VSSHQOO-----AAGNQSGSIGTORLNKNIKPTETPTPALIALKDRNTNL- 1421
Db 2070 IGTLOIKYDLHOTKYVTSIFAENATDNDNDLIE-----KEKEATQAI-----KULTNLF 2118
Qy 1422 ---SN-----YSKDIIMIKPKYLVERSIGVPMWSTGLDYGISQOTKDTGTSSSSQKGF 1471
Db 2119 TIDNSIDANLHNNKQW-----YFNSSELHKSISIKKLYKYM 2158
Qy 1472 DQDFQALGLKNTYHGLGLSIRIFDPGNELAKIDASKKGEELKLSYDLFKNYLNE 1531
Db 2159 HVFKLLNIGQINEKY-----FD-----ISKQFDDILOQESSELKENLND 2197
Qy 1532 YEKSPKIAKGTNIHPDQKEYPNPN-OQLPENY-----LNLVLNQPMKVTL 1577
Db 2198 LKKIQKISDKKKNFLNELSETPPNFNTLKEIYHIVKYESQIDEIKNITDENDNITL 2257
Qy 1578 YNSDPFITNLFVSPGSDRGSGTKLKQVIOQVNNNNYADMGSAYLTFWYDKNIIT--NQ 1634
Db 2258 Y--MDIITKLM-----KKVESIL-------NFVTYEND 2282
Qy 1635 PNVTANTADVPIDKVELEONTKLIAPNITQWPNISGSEKFKYKPTVFFGNWENENSS 1694
Db 2283 SNVIKOHITODNNDVSKIKNLKTIESFKILNKINEIKAQFY-----DNNN 2331
Qy 1695 MNSOQATPTWEKIR-----EGFALQALKSSFDQKTRTFVLTNAPLPLKYGPL 1743
Db 2332 INNVTISIQDVIDVKKHISKDLTIENELIHTQKLEDIKNSTYDIRSE---QIAKY--- 2385
Qy 1744 GPQNGPNFKTDWRLVLFQNDNQIAALRVQODRPEKSESDKQKWKFKVVIPEMFN 1803
Db 2386 -VNTIRNVQEQTKIQQNPNKDEIDDIQIIVNYNKESETK-----LPTITGN 2433
Qy 1804 SQNIRFVGVMQIQGPNLWLPVINSSVIYDFVRGFGSDNDVANLVNAPVQVKTIAFTNNA 1863
Db 2434 KNN-----DTSIISRINKFI--NLIKSEYNNNDNVSYNVAK---KLEEDVNI 2476
Qy 1864 FNNVFEFNISKKIVE 1879
Db 2477 ILDLKSNQWMLDLIQ 2492
RESULT 34
ID Q7REV3 PRELIMINARY; PRT; 2749 AA.
AC Q7REV3;
DT 01-MAR-2004 (TREMBLrel. 26, Created)
DT 01-MAR-2004 (TREMBLrel. 26, Last sequende update)
DT 01-MAR-2004 (TREMBLrel. 26, Last annotation update)
DE 235 kDa rhostry protein (Fragment).
GN Name=PY04930;
OS Plasmodium yoelii yoelii;
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=73239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=17XNL;
RX PubMed=12368865; DOI=10.1038/nature01099;
RA Carlson J.M., Angluoli S.V., Suh B.B., Kooij T.W., Perteu M.,
RA Silva J.C., Ermolaeva M.D., Allen J.E., Selengut J.D., Koo H.L.,
RA Peterson J.D., Pop M., Kosack D.S., Shumway M.F., Bidwell S.L.,
RA Shallom S.J., van Aken S.E., Riedmuller S.B., Feldblum T.V.,
RA Cho J.K., Quackenbush J., Sedegah M., Shoabi A., Cummings L.M.,
RA Florens L., Yates P.R., Raine J.D., Sinden R.E., Harris M.A.,
RA Cunningham D.A., Preiser P.R., Bergman L.W., Vaidya A.B.,
RA van Lin L.H., Janse C.J., Waters A.P., Smith H.O., White O.R.,
RA Salzberg S.L., Venter J.C., Fraser C.M., Hoffman S.L., Gardner M.J.,
RA Carucci D.J.;
RT "Genome sequence and comparative analysis of the model rodent malaria


```
Db 2257 SIDSIKOLYKQMHVFKLLNIGHINKY- - - - - FDISKEFDNLOQSE- - - - - 2300
QY 1520 KSYDLPKNYLNEYKSPKIAKGWTHIHDPQKEYPNP- - - - - QKLPENY- - - - - L 1565
Db 2301 - - - - - LKENLNDLKKIGOKISDKKKKFLNELSEIPNPENTLKEIYHEIVKYESQIDEIK 2355
QY 1566 NLVLNQPWKVTLNNSDPITNLVPEBEGSDRGSGTKLKQVIOQVNNYADWGSAYLTFW 1625
Db 2356 NITDENDNITLY- - - - - MDIITKLM- - - - - KKVESIL- - - - - 2383
QY 1626 YDQNIIT- - - - - NQPNVITANTADVIFIKVKELEDNTKLIAPNITQWNPENISGSKEKFKPT 1682
Db 2384 - - - - - NFVTYENDSNVIKQIHDNDNDVSKIDNKLKTTIESFOKILNKLEIKAQFY- - - - - 2437
QY 1683 VFFGNMENENSSMNSQAQPTWBEKI- - - - - REGFALQ- - - - - ALKSSPDQKT 1724
Db 2438 - - - - - DNNNNNNVISTISQVIDVKKHISKDLTIENELIEIQKSLEYIKKSTYDIRS 2489
QY 1725 RTFVLITNAPLPLWKYGLGFGQPNFKPTQDWLVLVFQNDN- - - - - QIAALRVOEQDRPEKSS 1782
Db 2490 EQITKYVN- - - - - PIHDY- - - - - VEQQTKKIQNDPNKDEIDDL-IOETVWYNKES 2533
QY 1783 EDKDKQKWKPKVPIPEEMNSGNIRFVGWQIQGPNLWLPVINS- - - - - SVIYDVPYGTGDS 1841
Db 2534 ELK- - - - - LPTIINKNKV- - - - - TPIISRIDKVINLIKSEYNN 2567
QY 1842 NDVANLNVA 1850
Db 2568 NDVSNVA 2576

RESULT 35
Q8WR56 PRELIMINARY; PRT; 2957 AA.
AC Q8WR56;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Normocyte-binding protein 1.
GN Name=NBPI;
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5633;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Dd2;
RX MEDLINE=21590493; PubMed=11733572;
RA Rayner J.C., Vargas-Serrato E., Huber C.S., Galinski M.R.,
RA Barnwell J.W.;
RT "A Plasmodium falciparum homologue of Plasmodium vivax reticulocyte
RT binding protein (PvRBP1) defines a trypsin-resistant erythrocyte
RT invasion pathway.";
RL J. Exp. Med. 194:1571-1581 (2001).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Dd2;
RA Rayner J.C., Barnwell J.W., Galinski M.R., Tran T., Huber C.S.,
RA Corredor V.;
RL Submitted (AUG-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF411931; AAL38220.2; -
DR ProDom; PD001963; Botulinum; 2.
SQ SEQUENCE 2957 AA; 356102 MW; 3BE03A234E78F52E CRC64;

Query Match 3.9%; Score 376.5; DB 2; Length 2957;
Best Local Similarity 18.6%; Pred. No. 1e-06;
Matches 352; Conservative 321; Mismatches 680; Indels 543; Gaps 88;

QY 50 IDSVAFKPSIANFT- - - - - SDYQSVKALLNG- - - - - KTFDPKSEFTDFVSKF 93
Db 771 IDDIKIETLQKITHIVNNIKTIKKDLKEFIQHLIKYMNERYQNMQQGYNNLTNYINQY 830
QY 94 DFLTNNGRTVLEIPKVVQVISEPSPEDDKERFLGHLKLELDG- - - - - NIAQSATKFI 149
```

```
Db 831 BEENNMQYITIRNIQKIYDNIYAKEKB-IRSGQYKDFITSRKNIYINREINISKNV 889
QY 150 YLLPLDMPKAAALGQYSYIVDKNFNNLIIPHLSNFSASQIKPLALTRSSDFIAKLQNFNQ 209
Db 890 DM1KNEKK- - - - - KIONCVDK-YNSEI- - - - - KQVVMKPKNGDTQD- - - - - ENNNNNN 931
QY 210 DELWVYLEKFPDLKANIQLQTADPSPFEKGNLVDVPVYSPFIRNPQCKEWASDLNQD- 268
Db 932 D- - - - - IYDKLIIVPLDSIKQINDKYNTHEHF- - - - - TTFINKIWNHKK- - - - - NQEM 974
QY 269 -OKTVRLRLTFEFSQAQKTIKADYKQDETFLSSSIDLKASNGTSLFANE-NDLKDQL- - - - - 323
Db 975 MEETVIVYKELK- - - - - ILK- - - - - ILNISLAKACEKNKNSINTLNDKTQELKKI 1018
QY 324 - - - - - DVDLDVSDYFGQSETITS- - - - - NSQKVPVPASERSLKDVRVFKDQOKP 370
Db 1019 VTHEIDLL- - - - - QKDILTQSISNKNVLLNLLKEI- - - - - EQYIIDHKLKKS- - - - - 1065
QY 371 RIEKESLEYDALSFSVQ- - - - - LOELVSKPN- - - - - SIKDLVNATLARNLFSLGKYNF 420
Db 1066 - - - - - DLFTYEQSKNYFYKKNKDNFDIQTINKONNEWLAKNYINEINKYQTYLYEKINV 1123
QY 421 LFDLLASHLDYVFLVSKAKIKOSSITKFLFIPLIKISLKSSILGDQEPNITLFEKYT 480
Db 1124 LLHNSKSYQVYFY- - - - - DHIINLIQKNYLENTLTKIQ- - - - - DNEHSLVALQNEEY 1174
QY 481 FKLDNFRD- - - - - VEIEKAFGLYPGVNBELEQ- - - - - RKQRASFEKE 520
Db 1175 QKVNEKQDQNEIKKIQLEKKNKNDILTYENNIEQKQIEQKIELKTNQAOKDQIVNTLNE 1234
QY 521 KSKKGLKPEFSQKESKAINN-QSG-LEEDDNITERLPENSPIOYQQENAGLQSPDKP 578
Db 1235 VKKKIITYYKVDNQISNLVKNYEGKVEYDKNVQNVNADDTDIDINDIDINDID 1294
QY 579 YM- - - - - IKDVQVQRYLAKSQIOELIKAKQYTKLAKLLSNRHTYINISL- - - - - LKEQLFDV 631
Db 1295 EINDIDEIKDIDHIKHFDTHFDIYHADD- - - - - TRDEYHIALSNYIKTELNI 1345
QY 632 NPR- - - - - IPSRDIKAKFVLDKTEK- - - - - NKYW- - - - - QIYSASP 665
Db 1346 NLQEIKNNIIKIFKGFSAHKEIKKESQIKNKFTKMDVIVINQLRDIRDQRMLDLYKE- - - - - 1402
QY 666 VQNKWSLFGYRYLLGLDPKQTIHELKVLQKAGLQPEGYENLPDQFNLEDLKNRIKT 725
Db 1403 -LDEKYSEFN- - - - - KTKIEINNIRENNINNVETLWYKNIIEYFLRYWQDKKA 1451
QY 726 PLFSQ- - - - - KDNFKLSLLDFN- - - - - NYDGEIKAPEGLPLFLPKELRRSS 769
Db 1452 AKYMNIDITYKNNIBIISKQINPENYVETLKTNNVSYVEKAND- - - - - LFY-KQINNII 1506
QY 770 NSGGQNSNSPW- - - - - BQEIISQFKQNLNQLAQFSTKIWEIKIGDENEFQ- - - - - NN 822
Db 1507 NS- - - - - NOLKNEAFTIDELQIQNRKNLLTKQOIIQYTNEI- - - - - ENIFNEIKNIN 1557
QY 823 RL- - - - - QYK-LLKDLQESWINKTRDNLVYTLVGLKLVKPKNNLEAKFROISNLOELLTAF 878
Db 1558 ILVLTNYKSILODISQ- - - - - NINHVSITYEQL- - - - - HNLVYKLEB- - - - - EKEQMTLY 1603
QY 879 YTSAAALSNWNYQDSGAKSTII-PEEIAELDPKVEKVGADVQYQKFIYAGFDDNAGK 937
Db 1604 HKSNNVLNQNINFNEDAFINLLINIEKIKNDITHIKEK- - - - - TNYIMDVNKS- - - - - KNAQL 1658
QY 938 FNOEIVRSSRTIYLTSGSKLEADTIDQANQAVNAPLQSGFYLDTFRFQVOKLAT 997
Db 1659 YFHNTLRGKEIEYLLKNLKNSTNQOITLQELQVQENVE- - - - - KVKDIYNQTIKY- - - - - 1709
QY 998 SLAVQHQKQKTLPPKLN- - - - - NDGYTLIHDLKPKVPVPOISSSEPKDFWFGKLNQOSQON 1055
Db 1710 - - - - - EEKIKGNTHIITDENKINDILHNSFIKQINME- - - - - SSNNKKQT 1750
QY 1056 - - - - - VNVSTFGSIIESPYFTNFQEDAD- - - - - LDQGDQD- - - - - DSRQGNLSLDN 1097
```

Db 1751 KQIDIIINDKTFEBHIKTSKTKINMLKEQSQKHIDKTLTLEQALKLFDVINSTNNLDN 1810
Qy 1098 QEAGLLKQKLAIIAGNQFIQYQNDKEIE--FEIINVEKVSLEL----- 1139
Db 1811 MLSEINS-----IQNNHTIYIQEANKSFDKFKIICDQVNDLNLKSLGLDLYNMHLKN 1864
Qy 1140 -----SFRVEFKL--AKTLEDNGKTRVLS-----DETWSLIIVNTTIETKPE 1179
Db 1865 LQNEIRNNMLKFNFLDKSKIDIEKKLIDILKVNISNNISLKLKYYEALFQVKKE 1924
Qy 1180 MSAPVEPDTKWVEQYDRTPLAAKTKVLAKFOQIPVDGS-----GN 1222
Db 1925 KAEIQENIEKIQINTLSVDFKPPFFI-----QLNTDSQHEKDINNVTETKNNIDE 1980
Qy 1223 ISDKWLASIPVIHOOMLRLSPVV-----KTIRELGKTEQOQQOQQOQQOQQOQ 1273
Db 1981 IYVFIQSYNLIQKYSSEIFSSTLNYIQTKEIKESIK-EQNQLNQEKEASVLLKNIKI 2039
Qy 1274 -----KAVRKBELETYNPKDEFNINPLTKAHLRULSLN-----VNNDPNYKIEDL 1320
Db 2040 NETIKLPQIKNERQNDVHNKEDYNLLQYLLNYMKNEMEQKKYKNDVHMDKNY----- 2094
Qy 1321 KVIKNEAGDQ-----LAFSLRANNIKRLMNTPTITADYNPPFYNNEDHRSIDKYLNN 1373
Db 2095 --VENNGEKEKLLKETISSYDKINNINNKJ-----YIYKKE-----DTVPNN 2137
Qy 1374 KGNVS-----SHQOQOAGNGQSGLIORLNKNIKPEFTPTPALIAKDRNNTNL 1421
Db 2138 MIKVSLEILLIILIKKQOQEQRVINAEYDSSLINK-DEEIKKE-INNOIIEI-LKHNENI 2194
Qy 1422 SNYSKIIIMPKYLVERSIGVPMWSTGLDGYIGSEQTKDGTSSSSQKGFQDFQIALGL 1481
Db 2195 SIIFKDIQIKKQ-----SQDIITNN-- 2215
Qy 1482 KNTYHGKLGISIRIFDQNE-----LAKIKDASNKGE--EKLLK----- 1520
Db 2216 --NDMHKSTILLVDIIQKKEALNKQKILRNIDNINLKNENIIDKVIKNCDDYKDILI 2273
Qy 1521 ---SYDLFKQVLYNEYEKKSPIAKGWTNIHPDQKEYPNPQKLPENY-----LNLVLN 1570
Db 2274 QNETEYQKQINHYEYKKSXI-----DILKIKIKQIKQIYKKNLEQMNTIIN 2324
Qy 1571 QPWKVTLYNSDFITNLVFEPEGSDRGSGTKLKQVIO-----KQVNNYADWGSAY-LT 1623
Db 2325 OSIEQHFVNADILQN-----BKILEEIIKRLDILDEQIMTYHNSIDELYKLG 2373
Qy 1624 FWYDKNIITNPQNVITANIADVFIKDVKELEDNTKL 1659
Db 2374 IQCDNHLITISVVVNRKTTKIMIHKKQKEDIQKI 2409
RESULT 36
Q81HY4
ID Q81HY4 PRELIMINARY; PRT: 2849 AA.
AC Q81HY4;
DT 01-MAR-2003 (TREMELrel. 23, Created)
DT 01-MAR-2003 (TREMELrel. 23, Last sequence update)
DT 01-MAR-2003 (TREMELrel. 23, Last annotation update)
DE Hypothetical protein.
GN ORFNames=PP11_0392;
OS Plasmodium falciparum (isolate 3D7).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=36329;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=22255705; PubMed=12369864; DOI=10.1038/nature01097;
RA Gardner M.J., Hall N., Fung E., White O., Berriman M., Hyman R.W.,
RA Carlson J.M., Pain A., Nelson K.E., Bowman S., Paulsen I.T., James K.,
RA Eissen J.A., Rutherford K., Salzberg S.L., Craig A., Kyes S.,
RA Chan M.S., Nene V., Shallow S.J., Suh B., Peterson J., Angiuoli S.,
RA Pertea M., Allen J., Selengut J., Haft D., Mather M.W., Vaidya A.B.,
RA Martin D.M., Fairlamb A.H., Fraunholz M.J., Roos D.S., Ralph S.A.,
RA McFadden G.I., Cummings L.M., Subramanian G.M., Mungall C.,

RA Venter J.C., Carucci D.J., Hoffman S.L., Newbold C., Davis R.W.,
RA Fraser C.M., Barrell B.,
RT "Genome sequence of the human malaria parasite Plasmodium
falciparum".
RL Nature 419:498-511(2002).
DR EMBL; AB014841; AA035975.1; -.
KW Hypothetical protein.
SQ SEQUENCE 2849 AA; 335917 MW; B5515D173D96B813 CRC64;
Query Match 3.8%; Score 373; DB 2; Length 2849;
Best Local Similarity 18.6%; Pred. No. 1.3e-06;
Matches 376; Conservative 332; Mismatches 725; Indels 592; Gaps 105;
Qy 82 KSEFTDFVSK-----PDLTNNRTVLEIPKYYQVVISFSPEDDKERPRGLPHLKEKL 136
Db 403 EDEKFNDMVKNKAKDINFEMWNSKKYSKIDINL---KYKDSNDKMKSN--DSHVK--- 455
Qy 137 EDGNIAQSATKFI--YLLPLDMPKAAALGOYSYIVD-----KNFNLIHLP----- 179
Db 456 SSENILNKNKILNRYDLPKDI-----SYVVEYQKRKEQAKNYHNVENPNYTEDK 507
Qy 180 ----LSNFAQSISIKPLALTRSSDFIAKLNQFNQDELWVY---LEKFFDLLEALKANIR 230
Db 508 NINDDKENYPGDNI-----NLNNIRSNVTYVYANSTLSNEKXNNLNSNIS 554
Qy 231 LQTADFSEKGNLVDPFVYSFIRNPQNKQEWASDLNODQKTV--RLYLRTFSPQAKTIL 288
Db 555 EDVSSIN--KENLLNTNI---NKSNETITQCCDNNKLVNNIYIDQEF-----LI 602
Qy 289 KDY-----KYKDETFLSSIDLKASNGTSLFANENDLKO-OLDVDLLDSDYFGOSSET 340
Db 603 KYNVNVKNKDKY--DSTYIDSL-----KXNNFIKDIYTDNDVIN--TESMGIYNET 649
Qy 341 ITSNSQVQKVPASERSLKDRVKF-----KKDQOKPRIEKS--LYE 379
Db 650 KXNKEKISE-PCIHNTSNIYEVILTNNDMTNVGNENSNDILNKNIEKTNFENESKKLYD 708
Qy 380 -YDALSFYSOQLQELVSKPNSIKD-LVNATLARNLRFSLGKYNFL---FDDLASHDY--- 431
Db 709 VYDMINDYKKNKEETINKIQEKCVDKVM-----YDFINNNIDKETNLDMDDKK 758
Qy 432 YFLVSKAKIKQSSITKKLFIEL---PIKISLSSILGDOEPNIKTLFEKEVTPLKNFR 487
Db 759 HLYLNNEKMNSTCDDKDTYLEMSKKEKINIFLYLKMIDVN-SLSHLFQ---YFVDEK 813
Qy 488 DVEIEKAFGLLYPGVNEEELQAKAQRASPEKSKKGLKEFSQOKEENSKAINNQBEGLE 547
Db 814 DEEMKKKLQFLIGGDE-----KKQMEFMEYKINQNTOTLNKGLKHESVOTNNEKKIQ 867
Qy 548 EDDNITERLPENSPIQY-QQENAGLGASPKPYMIKQVQRYLAKSQIQLKAKD-- 604
Db 868 IENIIQTDIKDITKTLYKNDMINKKTSIDSVF--FKSLSKDSYDLYNKN-KEDIKQNDTT 925
Qy 605 YTKLAKL-----LSNRHTYN-----ISLRKEQL---FDVNPRISSRSIEKAK 645
Db 926 YTKOEDMENKVDVTVNESNTETNEVOKINDLKVKILEKIKGCDYNVSNNNYNDEQTA 985
Qy 646 FVL--DKTE--KNKYQIYSSASPVFQNKWSLFGYYRYLLGLDPKQTIHELKVLGOKAGL 701
Db 986 ILMQLQDNKYSKEYKMDVYNL-----IDENRNI--LSKLNDEENM 1023
Qy 702 QFEGYEN-----LPDSFNLDELKNIIRIK---TPLESQKDNFKLSLLDF 741
Db 1024 KSNEYKKNQNRSMVTVETVFASLKSFEREMNLLKSHNERLRRIEKLYESDRIK----- 1077
Qy 742 NNYVDGRIKAPFGLPLFLPKELRRNSNGSGNSNSPWEQEIISQFKDQNLNQDQLA 801
Db 1078 NEYIKMB-KLKESQDRUFATE--RHIEKHLNEDLNLSKKNEDMKYDLKKNKI----- 1127
Qy 802 QFSTKIWEKIIGDENEPDQN-NRLQYKLLKDLQESWINKTRDNLNLYWTLGDKLVKPKQN 860
Db 1128 -----KIIALESQIDNNLNMISNNNEKDNNN---NNNNDN-----NNNNN 1164


```
QY 1107 LAI-LGN---QFIQYQONDKEIEFEIINVKVSELSFRVEFKLAKLTEDNGKTIRVLS 1162
Db 1917 SGIDLKFNINIAILPYLDSQKQDITLFTIPSEKPTSETYTKI-----SDSYNTLLDIL 1968
QY 1163 DETMSILVNTTIEKTPENSAVPEVDTKWEQYDPTPLAAKTKVLFKFKDQIIPVDSGN 1222
Db 1969 KRSQEL-----QKKEQALNLIFENRLHD-----KVQATNELKDTL--SDLKN 2010
QY 1223 ISDKWLASIPVIHQ--QMLRLS-----PVVKT-----IRELGKTEQOQOQ----- 1262
Db 2011 KKEQILNKVYLLHLSKLNELNKLSCNSQNYDTILESSKYDKIKKSNYNEKEKENLGINFD 2070
QY 1263 -QOQOQOQOQOQKAVKKEBELTYNPKDF-----NILNPLTKAHLRLTSLNVLNNDPN 1314
Db 2071 VRAMEQFNNDIKDIEKLENNYKSHKNYNFSEENNNILQSKKGLKELT--NAFNAETK 2128
QY 1315 YKIEDLVTKN-----EAGDHQAFSLRANNIKRLMNT--PITPADYNPFF-----YVN 1361
Db 2129 -KIEDKILBKNGLINLKIETRKDCMLFT-----YKTLVETLIKTYDTYKTFITSATKSK 2182
QY 1362 EDWRSIDKTLNKNKNVSSHQQAAGNQGSLIQRILNKNIKPETFTPALIALKDRNNTNL 1421
Db 2183 EPLKYIDATSNLNDINTLTQKYDLNQ-----INKHV-----ASMVADATNDNNN- 2228
QY 1422 SNYSOKIIMKPKYLVERSIGVPWSTGLDGYIGSEOTKDGTSSSSQQGFQDQFQALGL 1481
Db 2229 -----LIEKE-----KEATKT--INNLTFTIDSNNKIDADGL 2259
QY 1482 KNTYHGKLGSLTRIDPCNELAK-----IKDASNKKGEKLL-----KSYDLFQNY-- 1528
Db 2260 HNNK-----IQIIFYNSLHKSIDSIKQLYKQGHAFKLLNIGHINKKYFDISKEPDN 2311
QY 1529 -----LNEYEKSPKIAGWTNIHPDQKEYPNP--QKLPENY----- 1564
Db 2312 ILQOBSSELTAANDLKEIGQISDKKQFLHALSETPIPNFNTLKEIYVHDIVKYKNQID 2371
QY 1565 --LNLVLNQPKVTLNNSDFITNLVFEPEGSDRGSGTKLKQVIQOVANNVADWGSAYL 1622
Db 2372 EIEENITNEENITLY--MDII-----TKLMKKVQSILN-----EV 2405
QY 1623 TFWYDKNIITQNPVITANTADVFIKDVKELEDNTKLIAPNITOMWPNI SSGSEKFKYKPT 1682
Db 2406 TTY-----ENDSNIILKQIHDNDNDVSKIESLETTTQSFQKILNKLINEKAQFY--- 2456
QY 1683 VFFGNWENSSMNSQAQPTW-----EKIREGFALQALSKSPDQKTRTFVLTT 1731
Db 2457 -----DNNNNNVISTISQDVNDVKKHSKDLTIENELIQIKSLEDIKKSTYDIRS 2508
QY 1732 NAPLPLWKY-GPL-----GFQGNPNFKTQDWRLVFNQDNDNQIAALRVQODRPEKS 1781
Db 2509 E---QITKYVNPIDHYVVEQOTKKIQNNPN--KD-----EIDD-----LIQBIVNYNKE 2551
QY 1782 SEDKQKWKVKFVVPPEEMFNSGNIRFGVMQIQGPNTLWLPVINS-SVIYDFYRGTD 1840
Db 2552 SELK-----LPTIINNKNV-----TPIISRIDKVINLIKSEYN 2585
QY 1841 SNDVANLNAPWQVKTIAFTNNAFNVPKFNISKKIVE 1879
Db 2586 NNDNVSTNVAK--KLEEDANNIIRDLDTSNNMLNDLIQ 2621

RESULT 40
Q25662
ID Q25662 PRELIMINARY; PRT; 1939 AA.
AC Q25662;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Repeat organellar protein.
OS Plasmodium chabaudi.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5825;
```

```

[1]
RN SEQUENCE FROM N.A.
RP STRAIN=96V;
RA Werner E.B., Holder A.A., Aszodi A., Taylor W.R.;
RT "A novel 11-residue coiled-coil motif predicts a histidine zipper.";
RL protein 11p. Lett. 3:139-146(1996).
[2]
RN SEQUENCE FROM N.A.
RP STRAIN=96V;
RX MEDLINE=98418765; PubMed=9747969; DOI=10.1016/S0166-6851(98)00067-X;
RA Werner E.B., Taylor W.R., Holder A.A.;
RT "A plasmodium chabaudi protein contains a repetitive region with a
predicted spectrin-like structure.";
RL Mol. Biochem. Parasitol. 94:185-196(1998).
DR EMBL; U43145; AAC63403.1;
DR PIR; T18372; T18372.
DR HSSP; P13848; INO4.
SQ SEQUENCE 1939 AA; B36E462001C6F22F CRC64;

Query Match 3.8%; Score 370.5; DB 2; Length 1939;
Best local similarity 19.3%; Pred. No. 1e-06;
Matches 394; Conservative 325; Mismatches 683; Indels 643; Gaps 98;

QY 68 SVKALLN--GKTFDPKSEFTDFVSK-----PDFLTNNGRTVLEIPKXYQVVISFSS 118
Db 4 NLKSKKNEDSGNSKSKTNETSGIEKKEKSNKYKIVNNSTKKDKDNDSIVYDDES 63
QY 119 PEDDKERFLGFLHKE-KLEDGNTAQSATKTYLLPLDMPKAAALGOYSIVVDKFNENLLI 177
Db 64 KVGENDH-----HMKEYELEDO--LKETLKISITALSIKV-----EYEVKIEELEKELKL 111
QY 178 HPLSNFSAQSIKPLALTTRSSDFI-----AKLNQFNQDDELWV----- 214
Db 112 EKEQINKEYEK--ELNEKSEFIKQOMELLKELNINLNKENKINNKEIITLKREKLNLD 169
QY 215 ---YLEKFFOLEALK---ANIRLQ---TADFSPEKGNL--VDPFVYSFPIPNQOKEW 261
Db 170 ISEYIEKNEKEKLNVEVTNIKMSLDKLTCEVQEKDNLEKINKKVKIEKENNLRELKEF 229
QY 262 ASDLNQ-----DQTVRLYLRTFSPQAKTILKDYKYKDEFTFLSSIDLKASNGTS 311
Db 230 MKERNEIIESLDTGTTNDKKNAYEKLEISFEKRRMI----- 265
QY 312 LPANENDLKQDLVDLLDVSDYFGQSETITSNSOVKVPVASE---RSLKDRVKFKKQOQ 368
Db 266 ---EMLDSKLIKEENF-----ANKQAKLEKENEIIEKLD--IESREKDF 307
QY 369 KPRIEKSFLYEYDALSPYSQL-----QELVSPKNSIKDLVNTLARN 410
Db 308 KSKEEKFASMENELNTLKSLSKNACOMEVYKLEIKDLSQSLVEKEREIFEIKNE----- 362
QY 411 LRFSLGKYNFLFDDDLASHLDYFLVSKAKIKOSSITKKLFIELPIKISLSSILGQDEPN 470
Db 363 ---YDD-----KINNMKKELSSINDK-----GIDNTVLHSEBEK 393
QY 471 I-KTLFEKVTFKLONFRDVEIEKAFGLLYPGVNELEQAKAQRASPEKSKKKG----- 525
Db 394 INKLLKEKETELN-----EIHKKYNLEIHTIKNELNEKEE-----ELEKNKHAHTVE 440
QY 526 ---LKEFS--QOKEENSK-----AINNQEGL--EEDDNITERKLPENSPIYOQENA 569
Db 441 VTNLTKEIKLEKKTDEKAGECHKNELNELNQLSKLNKEKDNIT-----KNENT 488
QY 570 GLGASPDKPYMIKDQVONQRYYLAKSQIQELIKAK-----DYTKLAKLNSN-RHTYNIS-- 621
Db 489 EL-----NDKISSINSEVNTLNKDKQTLGNDIKTLNDLNLNKNENITSDN 534
QY 622 --LRLKEQLFDVNPRISSRDTEKAFVLDKTEKNKYWQIYSSASPVFQNKWSLFGYRY 679
Db 535 KNNKMKEDLAMLN-----EEMEGKCVWIDELEK-KY----- 564
QY 680 LLGLDPKQTIHEL-VKLGQKAGLQFEGYENLPSPDNLEDLKN-IRIKTFLPSQKDNF--- 734
```


GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 28, 2005, 20:54:27 ; Search time 60 Seconds
(without alignments)
3013.188 Million cell updates/sec

Title: US-10-607-631-8
Perfect score: 9732
Sequence: 1 MNKKSTLLATATAAIIIGST.....TNNAFNVKFNISKIVE 1879

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues
Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR 79:.*
1: pir1:.*
2: pir2:.*
3: pir3:.*
4: pir4:.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	394	4.0	1099	2 T18357	mhp1 protein - Myc
2	393	4.0	1092	2 T18354	adhesin - Mycoplas
3	391.5	4.0	1108	2 T18353	protein p97 - Myco
4	377	3.9	2269	2 T28677	rhostry protein -
5	370.5	3.8	1939	2 T18372	repeat organellar
6	366	3.8	2401	2 T28676	rhostry protein -
7	340.5	3.5	3394	2 T18501	hypothetical prote
8	333	3.4	2166	2 G70163	hypothetical prote
9	331.5	3.4	1738	2 T14867	interaptin - slime
10	325	3.3	4688	2 F82885	hypothetical prote
11	319.5	3.3	2819	2 A90551	conserved hypochet
12	312	3.2	2244	2 F90563	hypothetical prote
13	305.5	3.1	1875	2 S38173	myosin-like protei
14	301	3.1	1631	1 SAZQK1	major merozoite su
15	301	3.1	2663	1 S28261	centromere protei
16	295	3.0	1639	2 S05603	major merozoite su
17	291.5	3.0	1727	2 T50073	myosin-like coiled
18	291	3.0	1640	2 A24594	probable major sur
19	291	3.0	5005	2 F82884	hypothetical prote
20	290.5	3.0	1679	2 S48385	hypothetical prote
21	290.5	3.0	1979	2 C71622	hypothetical prote
22	289.5	3.0	1302	1 JC6009	surface-located me
23	287.5	3.0	3848	2 T17414	TipC protein - gli
24	287	2.9	1183	2 F90559	conserved hypochet
25	282.5	2.9	2829	2 A42771	reticulocyte-bindi
26	281	2.9	1805	1 A64224	hypothetical prote
27	281	2.9	2441	2 D71623	erythrocyte membra
28	280.5	2.9	1818	1 S73852	hypothetical prote
29	279.5	2.9	2231	2 D71870	hypothetical prote

30	279.5	2.9	3216	2 C90538	hypothetical prote
31	278	2.9	3724	2 T18427	hypothetical prote
32	276.5	2.8	1690	2 T13030	microtubule bindin
33	276	2.8	3259	1 A56539	giantin - human
34	274.5	2.8	1997	2 F71607	DNA helicase II BR
35	272	2.8	1516	2 B71619	RAD2 endonuclease
36	271	2.8	3225	2 I52300	giantin - human
37	270.5	2.8	1365	2 T30822	impl protein - Myc
38	270.5	2.8	2599	2 F90608	ABC transporter pe
39	270	2.8	1127	2 T28317	ORF MSV156 hypothe
40	267	2.7	1830	2 E82909	conserved hypochet
41	264.5	2.7	1088	2 T18559	hypothetical prote
42	264.5	2.7	1208	2 T39068	coiled coil protei
43	262.5	2.7	1701	2 A54498	major merozoite su
44	262.5	2.7	2295	2 B71621	probable membrane
45	261.5	2.7	2178	2 S55805	alpha-toxin - Clos

ALIGNMENTS

RESULT 1

T18357

mhp1 protein - Mycoplasma hyopneumoniae

C:Species: Mycoplasma hyopneumoniae

C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004

C:Accession: T18357

R:King, K.W.; Faulds, D.H.; Rosey, E.L.; Yandey Jr., R.J.

Vaccine 15, 25-35, 1996

A:Title: Characterization of the gene encoding Mhp1 from Mycoplasma hyopneumoniae and ex

A:Reference number: Z18890

A:Accession: T18357

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-1099 <KIN>

A:Cross-references: UNIPROT:Q49560; EMBL:U27294; NID:g1403588; PID:g1403589; PIDN:AAB034;

C:Genetic:

A:Genetic code: SGC3

A:Note: mhp1

Query Match 4.0%; Score 394; DB 2; Length 1099;

Best Local Similarity 20.5%; Pred. No. 3.8e-09;

Matches 261; Conservative 213; Mismatches 475; Indels 326; Gaps 59;

QY	1	MNKKSTLLATATAAIIIGSTVGVGLASKVKYGVNPTQGVISQLGLIDSVAPKPSIA	60
DB	1	MSKSKTFKIGTAGIVGVGLTVGLSSLAKYRSESPRKANDPAKVSTLAFSPYAF	60
QY	61	NFTSDYQSVKALLNGKTFDPKSSSEPTDFVSKPFDLTNNGRVLEI-----PKYQV	112
DB	61	ETDSDYKIVKRWLVDSNN-NIRNKE--KVIDSFPTKNGDQLEKINFQDPYTKAKITF	117
QY	113	VISEFSPDDKFRFLGPHLKEKLEDCGNTAQATKFIYLLPLDMPKAA---LCQYSYVD	169
DB	118	EILEIIPDDVNQNVKVFQALQKLNHGDIKSD---IYEQTVAFAKSNLLVAEFNFSLK	174
QY	170	KNFNNL-----IIHPLSNFS---AQSIKPLALTRSSDFTAKLNQFNNOELWVLEKFF	220
DB	175	KITEKLNQIENLSTKITNFADEKTSQKDPSTLRAIDFOYDLNTARNPEDLDIKLANYF	234
QY	221	DIEALKANI-RLQTADFSEKGNLVDPPVYGFIRNPNQKQEWASDLNQDKTVRLYLRT	279
DB	235	PV--LKNLNLNNAENKLPNNLGNIFKFGFAKD--SSTNQYVSIQNG---IPSLFKAD	288
QY	280	FSPQAKTILKDYKDET--FLSSIDLKASNGTSLFANENDLKQQLDVLDDVSDIFGGQ	337
DB	289	LSQSAREILLAS---PDEVQPVINILRMKKDNSSYFLNFDVNNL-----	331
QY	338	SETITSNQVQKVPASERSLKDVRKKDQKQRIEKFSLYEY-----DALSFY	386
DB	332	-----TLKN---MKEDLNKQNLNLSAYEFLADIKSGFFGDKRSSH	370
QY	387	--SQLQELVSKPNSIKDLVNATLARNLRFSLGKYNFLFDDL--ASHLDYVFLVSKAKIKQ	442

Db 371 TKSEISNLLNKENIYD-----FKYNGKFNDRLNSPNLEYSLSYDAASAQLDK 417
Qy 443 SSITKGLFIELPIKISLSSILGDQ--EPNIK-TLFEKEVTFKLDNFRDVEIEKAFGLLYP 500
Db 418 KD---KSIILPYRLEIKDKFFADDLYPDTDNILVKEGILKLTGFK-----KGPKIDLP 469
Qy 501 GVNELEOAKRAQASFEKSKKGLKEFSQOKEENSKAINNOEGLEDDDNITERLPENS 560
Db 470 NINQOI-----FKTEY---LPFFEKGEQAKL-----DYGNI----- 500
Qy 561 PIQYQENAGLGASDPKPYMIKDQVQRYYLAKSQIOELIKAKDYTKLAKLLSNRHTYNI 620
Db 501 -----NPNYNTQ-----LAKVEVEALFKGNKNQEIYQALDGNAYEF 536
Qy 621 SLRLKEQLFDVNPRI--PSSRDIEK-----AKFVLDTKKNKYQIYSS--ASP 665
Db 537 G-AFKSVLNSWTGKIQHPEKADIQRTHLQSVKIGNSVNLNQPTTKE-QVISLKSNN 594
Qy 666 VFQNKWSLFGYRYVLLGLDPKQTHIELVKLGQKAGLQFEGYENLPSDFNLEDLXNI----- 721
Db 595 FFKNGHOVASYPQDLYTKDCLTVLETLVDLAKWGLE--TNWAQPKG-AFYTKDIPAEA 652
Qy 722 -RIKTPLESQDNF-----KLSLDPNNYDGE--IKAPEFGLPLFLPKELRRNSNSGGSQ 775
Db 653 DKLAFLEWKKDPVYNQINIEIHLQSFNILARNDVIKSDGYGVLLLPQSVMTELEGKN--- 709
Qy 776 NSNSPWEQELISQFKDQNLNQDQLAQFSTKIWEKIIIGDENEFDQ--NNRLQYKLLKDLQ- 833
Db 710 -----EAOIFEALKYSLIEN---SAFKTTILDKNLLLEGDTFTFGDPLKAFKLAQF 760
Qy 834 ---ESWINKTRDNLWYTY-----LGDKLKVKPKNNLEAKFRQISN-LOELL----- 875
Db 761 NNFAPWA-KLDDNLQYSFEAIKKGTTKEGKREEVKKVKELDKNKIGILPQPPAAKPEA 819
Qy 876 ---TAFYTSAAALSNNWYQSGAKSTIIFRIBAE-----LDPKVEKVGCA- 918
Db 820 AKPVAAPKPVAA-----KPEAAKPVAAKPEAAKPVAAKPEAAKPVAAK 871
Qy 919 -----DVYQLKFAVIAFGDDNAGFNQOEIVIRSSRTIYLTSG 956
Db 872 PVATNTNTWGTSLNKNPKEDYFPMAFSVKLEYTDE-----NKLSTPTPINVLELVH 927
Qy 957 KSKLE-ADTIDOLNQAQKAPLQSFYLDTERFGVQFKLATSALVQHKQKKEKTLPKKLN 1015
Db 928 QSEYEDQKIIKELDKTVLNLQYQFQVKTSDQ---YOKLSHPMMTEGTQNGKKGEGTP 984
Qy 1016 NDGYTLIHDKLKPKVPIQISSSPKDWPEGLNQ---NGQSNVNVSTF-----GSI 1064
Db 985 NQG-----KKAEGAPSQGKABGTPNQGKABGTPNQGKABGTPNQGKABGTPN 1037
Qy 1065 IESPVFTSNFQDADLDQD-----QDSSQGNNSLDNQEAGLLKQKLAILLG 1112
Db 1038 IKKO--GKNWTEVELIEDNLAGDAKLLYFLTRDDSKG-----DPKSSLKVKIITVQS 1090
Qy 1113 NQFTQYQONDKEIE 1127
Db 1091 N-----NNQELK 1097

RESULT 2

T18354
adhesin - Mycoplasma hyopneumoniae
C:Species: Mycoplasma hyopneumoniae
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004
C:Accession: T18354
R:Wilton, J.L.; Scarnan, A.L.; Walker, M.J.; Djordjevic, S.P.
Microbiology 144, 1931-1943, 1998
A:Title: Reiterated repeat region variability in the ciliary adhesin gene of Mycoplasma
A:Reference number: T18887; MUID:98361039; PMID:9695926
A:Accession: T18354
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA

A;Residues: 1-1092 <WIL>
A;Cross-references: UNIPROT:O07132; EMBL:AF001398; NID:g2150123; PID:g2150124; PIDN:AAC35
C;Genetics:
A;Genetic code: SGC3

Query Match 4.0%; Score 393; DB 2; Length 1092;
Best Local Similarity 20.6%; Pred. No. 4.2e-09;
Matches 261; Conservative 204; Mismatches 464; Indels 340; Gaps 59;

Qy 1 MNKKSTLLATAAIIGSTVFGTVGLASKVKYGVNPTQGVISQLGLDSVAPKPSIA 60
Db 1 MSKSKTKFKLGLTAGIVGLGVGLTVGLSLAKTRSRSPRIANDFAAKVSTLAFSPYAF 60
Qy 61 NFTSDYQSVKALINGKTFDPKSEFTDFVSKFDFLTNNGRTVLEI-----PKYQV 112
Db 61 ETDSDYKIVKRWLVDN--NIRNKE--KVIDSFSTFKNGDQLEKINFQDPPEYTKAKITF 117
Qy 113 VISEFSPEDDKERPRGLGFHLKEKLEGNIAQSAATKYIYLLPLDMPKAA---LGQYSIVD 169
Db 118 EILEIIPDDVQNQPKVAFQALQKLHNGDIAKSD---IYEQTAPAKQSNLLVABFNFSLK 174
Qy 170 KNFNNL-----IHLPLSNFS---AQSIKPLALTRSSDFTAKLNQFNQNODELWYLEKFF 220
Db 175 KITEKLNQNIENJSTKITNFADKTSQKDPSTLRAIDFQYDLNTARNAEDELDTKLANYP 234
Qy 221 DLBAALKANI-RLQTAQDFSEKGNLVDPPVYSFIRNPQNKQEWASDLNQDQKTVRLYLRT 279
Db 235 PV--LKNLINLNAPENKLPNNLGNIFEFSFAKD--SSTNQYVSIQNG---IPSLFLKAD 288
Qy 280 FSOQAKTILKDYKYKQBT--FLSIDLKASNGTSLPANENDLKDQDVLDDVSDYFGGQ 337
Db 289 LSQAREILAS---PDEVQVVINILRLMKDKNSSYFLNFDVNNL----- 331
Qy 338 SETITNSQVQKVPASERSLKDVKPKQDQKPRIEKFSLYEY-----DALSFY 386
Db 332 -----TLKN---MOKEDLNKAGQNLSEYFLADTKSGFFQDKKSSH 370
Qy 387 --SOLQELVSKPNSIKDLVNNATLARNLRFSLGKYNFLFDDL--ASHLDYFLVSKAKIKQ 442
Db 371 TKAISNLLNKENIYD-----FGKYNKFNDRLNSPNLEYSLSDAASASLDK 417
Qy 443 SSITKGLFIELPIKISLSSILGDQ--EPNIK-TLFEKEVTFKLDNFRDVEIEKAFGLLYP 500
Db 418 KD---KSIILPYRLEIKDKFFADDLYPDTDNILVKEGILKLTGFK-----KGPKIDLP 469
Qy 501 GVNELEOAKRAQASFEKSKKGLKEFSQOKEENSKAINNOEGLEDDDNITERLPENS 560
Db 470 NINQOI-----FKTEY---LPFFEKGEQAKL-----DYGNI----- 500
Qy 561 PIQYQENAGLGASDPKPYMIKDQVQRYYLAKSQIOELIKAKDYTKLAKLLSNRHTYNI 620
Db 501 -----NPNYNTQ-----LAKVEVEALFKGNKNQEIYQALDGNAYEF 536
Qy 621 SLRLKEQLFDVNPRI--PSSRDIEK-----AKFVLDTKKNKYQIYSS--ASP 665
Db 537 G-AFKSVLNSWTGKIQHPEKADIQRTHLQSVKIGNSVNLNQPTTKE-QVISLKSNN 594
Qy 666 VFQNKWSLFGYRYVLLGLDPKQTHIELVKLGQKAGLQFEGYENLPSDFNLEDLXNI 725
Db 595 FFKNGHOVASYPQDLYTKDCLTVLETLVDLAKWGLE-----ETNRAQFPKVEFYTK--- 646
Qy 726 PLFSQDNFK-----LSLDPNNYDGE--IKAPEFGLPLFLPKELRRNS 769
Db 647 DIFAEADKLFLGKGGKDPVYNQIKETIHLQSFNILARNDVIKSDGYGVLLLPQSVMTELE 706
Qy 770 NSGGSQNSPWEQELISQFKDQNLNQDQLAQFSTKIWEKIIIGDENEFDQ--NNRLQYK 828
Db 707 GKN-----EAOIFEALKYSLIEN---SAFKTTILDKNLLLEGDTFTFGDPLKAF 754
Qy 829 LKDLQ-----ESWINKTRDNLWYTY-----LGDKLKVKPKNNLEAKFRQISN-LOELLTAFY 879
Db 755 LKAAQFNNFAPWA-KLDDNLQYSFEAIKKGTTKEGKREEVKKVKELDKNKIGILP--Q 811

```
QY 880 TSAALSNNWYQDSGAKSTIIIEETAEILD-----PKVKEKYGA----- 918
D 812 PPAA-----KPEAAKPVAAKPEAAKPEAAKPEAAKPVAAKPVAAKPVAAKPVATNTN 863
QY 919 -----DVQLAFHYAIGFDDNAGFNQEVIRSSRTIYLKTSKSKLEAD-T 964
D 864 TWGFSLTWKPREDYFPMAFSKLEYTDE-----NKLSTKPTPEINVLVHQSVEEQKI 919
QY 965 IDOLNOAVKNAVLGLQSFYLDTERFGVFOKLATSLAVQHKKQKEKTLPKKLNNDGYTLIHD 1024
D 920 IKELDKTVLNLOVQFQEVKVTSEQ-----YOKLSHPMMTEGSPNGKKAEGAPNQ----- 970
QY 1025 KKKPV-----IPQISSPEKDMFEGKLNQ-----NGOSQNVNVSTF-----G 1062
D 971 --KKAEGAPSQGKKAEGAPNQKKAEGSPQKKAEGASNQOSTTTELTYNLPGLKKID 1028
QY 1063 SIIESPVSTNFQEDADLDQD-----QDSSRQGN-----NSLDN 1097
D 1029 EIIKKQ--GKNWKEVELIEDNIAGDAKLLYFVLRRDSSKSGDPKSSVKVKITVKSQNN 1086
QY 1098 QEAGLLKQK 1106
D 1087 QE---LKS 1092

RESULT 3
T18353
C:Species: Mycoplasma hyopneumoniae
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004
C:Accession: T18353; T03805
R:Hsu, T.; Artushin, S.; Minion, F.C.
J. Bacteriol. 179, 1317-1323, 1997
A:Title: Cloning and functional analysis of the P97 swine ciliu adhesion gene of Mycoplasma hyopneumoniae
A:Reference number: Z18886; MUID:97175562; PMID:9023217
A:Accession: T18353
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1108 <HSU>
A:Cross-references: UNIPROT:Q49542; EMBL:U50901; NID:g1399524; PID:g1399525; PIDN:AAB478
R:Hsu, T.; Minion, F.C.
Gene 214, 13-23, 1998
A:Title: Molecular analysis of the P97 ciliu adhesion operon of Mycoplasma hyopneumoniae
A:Reference number: Z15097; MUID:98391007; PMID:9729120
A:Accession: T03805
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1089-1108 <HS2>
A:Cross-references: EMBL:AF012905; NID:g2654768; PIDN:AAC32526.1; PID:g2654769
A:Experimental source: strain 232A
C:Genetics:
A:Genetic code: SGC3
A>Note: P97

Query Match 4.08; Score 391.5; DB 2; Length 1108;
Best Local Similarity 20.24; Pred. No. 5e-09;
Matches 258; Conservative 210; Mismatches 486; Indels 325; Gaps 57;

QY 1 MNKKSTLLATAAIIGSTVFTVGLSKVKYRGVNPQTGVISQLGLDSDVAFKPSIA 60
D 1 MSKSKTFKGLTAGIVGLGVGLTVGLSSLAKVRSSEPRKIANDFAAKVSTLAFSPYAF 60
QY 61 NFTSDYOSVKKALNGKTFPPKSSSEFTDFVSKDFLTNNGRTVLEI-----PKYQV 112
D 61 ETDSDYKVKRWLVDSNN-NIRNKE--KVIDSPSFFTKNGDQLEKINFQDPPEYTKAITF 117
QY 113 VISFSPEDDKERFLGHLKLEKLEGNIAQSATKTYLLPLDMPKAA---LQYYSIVD 169
D 118 EILEIIPDDVNQNFVKVQFQALQHLNGDIKSD---IYQTVAFAPAKOSNLLVAEFNLSK 174
QY 170 KNFNNL-----ITHPLSNFS---AQGIKPLATRTSSDFTAKLNQFNQDELWYLEKFF 220
D 175 KITEKLNQOENLSTKINTFADEKTSQKQSPSTLRAIDFQYDLNTARNPEDLDIKLANYP 234
```

```
QY 221 DLEALKANI-RLQADPFSEKGNLVDPPVYGFIRPNQONKEWASDLNODQKTVRLYLRT 279
D 235 PV--LKNLNLNNAPEKNLNNLGNIFEFAKD-SSTNQVVSIONQ---IPSLFLKAD 288
QY 280 FSPQAKTILDKYKDEET--FLSSIDLKASNGTSLFANENDLKDQDLVDLLDVSYFGQ 337
D 289 LSQSAREILAS--PDEVQPVNILLRLMKONSSYFLNPFEDFVNNL----- 331
QY 338 SETTITSNQKVPVPAERSLKDQVFKKQKQKQKPRIEKPSLEY-----DALSPY 386
D 332 -----TLKN---MQKEDLNKQONLSAYEFADIKSGFPFGPKRSSH 370
QY 387 --SQQLVSKPNSIKDLVNATLARNLRFSLGKYNFLPDDL--ASHLDYFVLVSKAKIKQ 442
D 371 TWAEISNLNKENIYD-----PKYNGKFNDRLNSPNLEISLDAASASLDK 417
QY 443 SSITKKLFIPIKISLSSILGDO-EPNIK-TLFEKEVTFKLDNPRDVEIEKAFGLLYP 500
D 418 KD---KSIVLIPYRLEIKDPFADDLYPDTKONILVKEGILKLTGFK-----KGSKIDLP 469
QY 501 GVNBELEQARKAQRASFKEKSKGKLBPSQOKSEKAINNOEGLEEDDNITERLPENS 560
D 470 NINQOI-----FKTEY---LPFFPKGKBEQAKL-----DYGNIL----- 500
QY 561 PIYOQENAGLGASDPKPYMIKDVQNRVYLAKSOIQELIKAKDYTKLAKLLSNRHTYNI 620
D 501 -----NPYNTQ-----LAKVEEALFKGNKQOEIYQALDGNVAYEF 536
QY 621 SLRLKEQLFDVNPRI--PSSRDIEK-----AKPVLDTKKNKQWQIYSS--ASP 665
D 537 G-AFQSVLNSWTGKIQHPEKADIQRTRHLBQVKIGSNVLNQPOPTTKE-QVSSILKSN 594
QY 666 VFQNKWSLFGYRYLLGLDLPKQTIHELKLDQKAGL-----QF-SGYENLPDGF--NLED 717
D 595 PFKNGHQVASYFQDLTKDKLTILETLVYDLAKWGLETNRAQFPKGVFOYTKDIFAEADK 654
QY 718 LKNIRIKTLPFSQDNF---KLSLLDFNNYDGB--IKAPEFGLPLFLPKELRRSSNSG 772
D 655 LKPLEUK-----KQDPYQIKHISLNFILARNVDIKSDGFVGVLLQLSQSVKTELEGK 709
QY 773 GSONSNSPWEQIISQFKDONLSNODQLAQFSTKIWEKIIGDENEFDQ--NNRLQYKLLKD 831
D 710 -----EAQIFEALKKYSLIEN---SAPKTTILDKNLLEGTDFKTFGDFLKAFFLKA 757
QY 832 LQ-----ESWINTKTRNLTYTY-----LGDKLKVKPKNNLEAKFRQISN----- 870
D 758 AQFNFPAPWA-KLDDNLOYSFEAIKKGGETTREGKREEDVKVKELDNKIKGILPQPPAAK 816
QY 871 ---LQELLTAFYTSAAALSNNWYQDSGAKSTIIPEEIAE-----LDPKVKE 914
D 817 PEAAKPVAAKPETTKPVAAKPEAAKPVAAKPEAAKPVAAKPEAAKPVAAKPEAAK 876
QY 915 KVGA-----DVYQKUFHYAIGFDDNAGFNQEVIRSSRTIYL 952
D 877 PVAAPKPEAAKPVATNTGFSLTNKPREDYFPMAFSKLEYTDE---NKLSTKPTPEINVL 932
QY 953 KTSKSKL-EADTIDOLNOAVKNAVLGLQSFYLDTERFGVFOKLATSLAVQHKKQKEKTL 1011
D 933 ELVHQSEYEBEQEIIKELDKTVLNLOVQFQEVKVTSDQ---YOKLSHPMMTEGSSNOGKS 989
QY 1012 KKLNDGYTLIHDKLLKPKVPIPOISSPEKDMFEGKLNQGNQSQNVNVS-----T 1060
D 990 EGTPNQ-----KKAEGAPNQKKAEGTTPNGKKAEGAPSQOSTTTELTYNLPDLGKK 1042
QY 1061 FGSIIESPYSTNFQEDADLDQD-----QDSSRQGNNSLDNOEAGLLKQKLA 1108
D 1043 IDEIHKQ--GKNWKEVELIEDNIAGDAKLLYFVLRRDSSKSG-----DPKSSSLVKIT 1095
QY 1109 ILLGNQFIQYQNDKGE 1127
D 1096 VKQSN-----NNOEPE 1106
```

RESULT 4

T28677
rhoptry protein - Plasmodium yoelii
C:Species: Plasmodium yoelii
C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004
C:Accession: T28677; C45521
R:Keen, J.; Sinha, K.; Brown, K.; Holder, A.
Mol. Biochem. Parasitol. 65, 171-177, 1994.
A>Title: A gene coding for a high molecular mass rhoptry protein of Plasmodium yoelii.
A:Reference number: Z20508; MUID:95021522; PMID:7935623
A:Accession: T28677
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-2269 <KE>
A:Cross-references: UNIPROT:Q26223; EMBL:L27838; NID:g457145; PID:g457146; PIDN:AAA21304
R:Keen, J.; Holder, A.; Playfair, J.; Lockyer, M.; Lewis, A.
Mol. Biochem. Parasitol. 42, 241-246, 1990
A>Title: Identification of the gene for a Plasmodium yoelii rhoptry protein. Multiple cd
A:Reference number: A45521; MUID:91101660; PMID:2270106
A:Accession: C45521
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 2131-2269 <KE2>
A:Cross-references: GB:M34283

Query Match	3.9%	Score 377;	DB 2;	Length 2269;
Best Local Similarity	18.5%;	Pred. No. 5.6e-08;		
Matches	418;	Conservative 363;	Mismatches 762;	Indels 712;
	Gaps 106;			
Qy	19	STVFGTVGLASKVKYRG----	VNPQGVISQLGLDSVAFKPSIANFTSDYQSVKKAL	73
Db	165	NTIYNTIKSVSDQI--YGGDIDTF	FNELSSIVKE--DPIDDIEDTKLENLSKIDNVYDKI	222
Qy	74	LNKTFDPKSEFTDFVSKDFLTNNGR----	TVLEIPKPYQVVVISFSPS-----	120
Db	223	-----QKMEIETVSKHLNNTIETN	KNLPTNLTLEI--KKY--IVDEISKELNKLWEDFXN	271
Qy	121	-----DDKERPRLGPHLKEKLED	NGIAQSATKFYLLPLDMPKAAALQCYSVYVD	169
Db	272	KEKELSNKISDYDKRQOLSEYKSM	LEIRHNYSQTN-----VONTKEEAAKQNY--D	323
Qy	170	KNFNLLIHLNFSNFAOSIKPLALTR	SSDRIAKLN-----QFNNOELWVYLE--KFFDL--E	223
Db	324	KSNEHMTTIPITNEDEISKIISEV	KTMKDEILSKVNTYIDFNKKYKETVNSEHQSQTELT	393
Qy	224	ALKANI---RLQADPFSEKGNLVDP	PFVYFIRNPQN--QKEWASDLNQDQTVRLYLRL--	277
Db	384	KIRAEVSDKELKKEQSFQDNOK---	SLINETKNSIEKEY-QNIN-TLKVKDYEVYK	434
Qy	278	-----TFSPQAKTILKD-----	YKYKDETFLSSIDLKASNGTS	311
Db	435	CKSTKESITFESSK-QTILKMLN	QNIKTVYKETSIDKSYIEKFEQILITGKQTKLENKFT	493
Qy	312	LF-----ANENDLKOQLDVL	LDVSDYFGQSEITTNSQVVPVPAERSLKDQVKKPK	365
Db	494	EFSLNHNHEANNELIKYFS----	DLKANLGINENMLYNQFTE-----KEKTFND--IKEN	544
Qy	366	DOQKPRTEKFSLYEDALSPVSO	LOELVSPKNSIKDLVNATLARNLFSLGKYNFLFDDL	425
Db	545	IHNELSKIEIKIH--ASIYNTSE	ETERIGINISLNTKVFYKENVTNLNIKEKL	602
Qy	426	ASHLDYVFLVSKAKIKOSSITKKL---	FIELPIKISLKSSILGDQBPNIKTLPF--KEVT	480
Db	603	-KHDFDSDFGKEGNIKYITDIK	INDIMAVSQIQDQHINGLDDIQKGSSEYVSEKQEI	661
Qy	481	FKLDNFRDVEI--EKAFGL-----	LYPGVNEELEQARQAQASFEKKS	522
Db	662	NKLEKVNTSEISNDNVGIIKKKQ	QIIVTKIDKKXNIYEEINKLSELSKTEKONTSLKV	721
Qy	523	KK-----GLKEFSQKKEENS	KAINNOGLEED--DNITERLPE-----NSPI	562
Db	722	KDNLISYGONLGNLFLROI	DEEKKKAENTIKSMEAYIDDLDNIIKKKSQEIETEMDKIWDI	781

Qy	563	QYQOENAGL	GAS	PD	KPM	IK	DV	Q	N	O	R	-----	V	L	A	K	S	O	I	O	E	L	I	K	A	D	Y	-----	F	K	L	A	K	L	S	613																					
Db	782	NKMEAL	K	I	SH	D	D	K	K	C	H	D	K	S	K	N	H	E	N	I	S	I	D	Y	K	S	K	I	O	D	F	R	E	S	D	I	N	D	I	N	K	N	K	L	Q	N	V	S	841								
Qy	614	NRHTY	N	I	S	L	R	L	K	E	Q	L	F	D	V	N	R	I	P	S	S	R	D	I	E	K	A	K	F	V	L	D	K	T	E	K	N	K	Y	W	I	Y	S	S	A	S	P	V	F	Q	N	K	S	L	673		
Db	842	ESQNH	S	D	I	---	N	O	C	L	N	E	V	A	N	I	N	I	L	K	N	I	K	I	I	D	K	V	K	E	-----	Y	T	S	E	I	E	K	N	-----	886																
Qy	674	FGYRY	L	L	G	L	D	P	K	O	T	H	E	L	V	K	L	G	K	A	G	L	O	P	E	G	Y	E	N	I	P	S	-----	D	F	N	L	E	D	-----	L	K	N	I	R	722											
Db	887	-----	K	N	I	N	D	E	L	N	N	E	K	I	K	E	G	L	S	K	E	C	R	S	K	I	N	S	T	L	D	D	X	I	D	E	C	I	K	N	I	934															
Qy	723	I	-	K	T	P	L	F	S	O	K	D	N	F	-----	K	L	S	L	D	F	K	N	-----	Y	Y	D	G	E	I	K	A	-----	P	E	F	G	L	756																		
Db	935	V	L	K	N	I	L	E	E	T	N	I	N	H	P	K	N	A	E	E	N	K	I	V	L	S	N	F	N	I	E	M	A	D	N	K	S	O	I	L	E	K	K	O	N	G	T	N	D	H	Y	N	I	994			
Qy	757	P	L	F	L	P	K	E	L	R	N	S	N	S	G	-----	S	O	N	S	N	S	-	P	W	E	Q	E	I	S	Q	F	K	D	N	L	S	N	O	Q	L	A	O	P	S	T	K	I	W	E	K	810					
Db	995	-----	K	E	L	S	H	K	D	K	S	G	Y	K	T	E	A	D	Q	N	K	A	I	O	K	N	K	E	L	F	E	Q	Y	K	E	E	-----	V	T	L	L	N	K	1038													
Qy	811	I	I	G	D	E	-	N	E	D	Q	N	R	L	O	Y	K	L	L	D	L	O	E	S	W	-----	I	N	K	T	-----	I	N	K	T	840																					
Db	1039	Y	A	V	E	L	N	K	N	E	D	K	T	N	D	S	K	O	I	I	K	E	I	K	O	A	H	N	Y	T	L	E	S	G	K	E	K	O	M	E	I	K	N	E	I	H	I	E	D	E	V	A	N	D	1098		
Qy	841	R	D	N	L	Y	T	M	L	G	D	K	L	K	V	P	-----	K	N	L	E	A	K	F	R	Q	I	S	N	L	-----	871																									
Db	1099	K	S	N	K	A	I	T	S	I	-	K	V	S	V	E	P	F	K	T	I	K	I	N	E	I	R	T	K	S	D	D	C	L	K	E	T	N	D	L	E	-----	K	O	I	S	N	L	S	I	D	T	Q	E	T	K	1153
Qy	872	-----	Q	E	L	L	T	A	F	T	S	A	L	S	N	N	Y	O	D	S	G	A	K	T	I	I	F	E	E	I	A	E	L	D	P	V	K	E	K	V	G	A	D	V	Y	Q	L	K	F	H	926						
Db	1154	L	T	E	N	G	K	O	L	T	B	E	L	L	E	S	L	K	K	O	K	N	I	E	D	O	K	-----	K	E	L	D	E	V	N	S	K	I	K	-	N	T	E	N	T	O	N	H	O	K	N	1205					
Qy	927																																																								

1394 IORLNKNIKPTFTPALIAALK-----DRNVTLSN-YSDKIIMIKPKYILVERSIGVPMST 1447
1729 IE-----KEKATOAKINLTTLFTIDSNNIDANALHNNKIOMV----- 1766
1448 GLDGYIGSOTKDGTSSSQKGFQDPTQALGLKNTYHGLGLSIRIFDPGNELAKIK 1507
1767 -----YFNSLHKSIESIKOLYKMHVFLNLTNGQNGKY-----FDLSKQFDNLL 1812
1508 DASNNKGBEKLKSLVDLFPKYNLYNEKSPKIAKAGWNTNHPDQKYPNPN-OKLPENYLN 1566
1813 Q-----LOESBLTAN-LNDLKEIGOKISDKKNKPLHALNETPIPNFNTLKEIYHD 1861
1567 LV-----LNQPKWTLVNSDDFTNLVPEPEGSGRSGTKLQVIOKQVNNYADWGSAY 1621
1862 IVKYRQIDIEINITSSEENI--TLIID-----TITKLKEKVOISILN-----F 1903
1622 LTFWYDKNITNOPNVTIANIADVFVKDVELEDNNTKLAPNITOWMPNISGSEKFKYK 1681
1904 VTY-----ENDSNIIKQIOTDNENDVSKIKESLKTIIQSFOELNKLNGIKAKQFYDN 1957
1682 TVFFGNWENESSMNSQAOTPTWEKIR-----EGFALQALKSSPDQKTRTFVLTTNAP 1734
1958 ----NNINNTISTISQDVNDVKKHISKOLTIENELIEIQSLSDIKNSTVEI----- 2008
1735 LPLWKYGLPGFQNGPFKTDQRLVLFQNDNDQIAALR--VQSDRPEKSSDKQKQWIK 1792
2009 -----RGAN-----NNYVNTIRNVEOQTNKIQNSNKDEIDDI 2043
1793 FKVVIPFEMFNSGNIRFVGVMQIQGNTLWLVPINSSVIVDFYRGTDGSDNDVANLVAWP 1852
2044 QK-----ILNKNKSETKLPYI-----TGKNKNTSI----- 2070
1853 QVKTIATFNAPNVPKE-----FNISKKIVE 1879
2071 ----ISKINKVINLIESEYGNNNVSYNVAKLEE 2101
RESULT 5
T18372
repeat organellar protein - Plasmodium chabaudi
C:Species: Plasmodium chabaudi
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004
C:Accession: T18372
R:Werner, E.B.; Taylor, W.R.; Holder, A.A.
Mol. Biochem. Parasitol. 94, 185-196, 1998
A:Title: A Plasmodium chabaudi protein contains a repetitive region with a predicted spe
A:Reference number: Z18922; MUID:98418765; PMID:9747969
A:Accession: T18372
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1939 <WER>
A:Cross-references: UNIPROT:Q25662; EMBL:U43145; MID:g1151157; PID:g1151158; PIDN:AAC634

Query Match 3.88; Score 370.5; DB 2; Length 1939;
Best Local Similarity 19.3%; Pred. No. 8.5e-08;
Matches 394; Conservative 325; Mismatches 683; Indels 643; Gaps 98;
QY 68 SVKALIN--GKTPDPKSSSFTDFVSK-----FDLTNNGRTVLEIPKKYQVVISFSS 118
DB 4 NLKSKNGDSNKDSKKTNETSGIEKKESKNWYKIVNNTKKDKKNDISIVYDDES 63
QY 119 PEDDKERFRIGFHLKE-KLEDGNIAQATFIYLLPLDMPKALQVSYIVDKNFNLLII 177
DB 64 KVGENDH----HMKEVELEDQ--LKETLSITALSIVK-----EYEVKIELEKELKL 111
QY 178 HPLSNFSAQSIKPLATRSDFI-----AKLNQFNQDELWV----- 214
DB 112 EKEQINKNEYEK--ELNKESEFTKQRMELLKEKELNLTNKENKINNKEITTLKREKIND 169
QY 215 ----YLEKFDLEALK---ANIRLQ----TADFSFEKGNL--VDPFYYSFIRNPQKQEW 261
DB 170 ISEYIEENKEKEKLYEVNTIKMSLDKLTCVEQEKDKNLEKINKKVIENKNNLRELKEF 229

QY 262 ASDLNO-----DOKTVRLYLRTFBSFOAKTILDKYKYKDETFLLSSIDLKASNGTS 311
DB 230 MKENNEIIESLDGINDKNAYEKLKLEISFEKRM----- 265
QY 312 LPAENDLKDQDLVDLLSDYFGQSSETITSNSQVKVPVASE-----RSLKORVKPKKQOQ 368
DB 266 -----EMLDSKLEKENF-----ANKQAKLEKENEIIIEKLD-IESREKDF 307
QY 369 KPRTEKPSLYEYDALSPYSOL-----OELVSKPNSIKDLVATNATLARN 410
DB 308 KSKEEFASMENELMTLKSJLKNACQMEVYKLEIKOLDSLVEKERELEIFEKNE----- 362
QY 411 LRPSLGKTNFLFDDLASHLDYVFLVSKAKIKOSSITKKLFIELPIKISLKSSILGDQEPN 470
DB 363 -----YDD-----KINNKEKLSLNDK-----GIDNTVLHSEEBK 393
QY 471 I-KTLFEKEVTFKLDNFRDVEIEKAFGLLYPGVNBELQARKAQASPEKESKKG----- 525
DB 394 INKLKKEKETELN-----BIHKYKYLEIETIKNELNEKEE-----ELEKKAHTVE 440
QY 526 ----LKEFS--OQKEENSK-----AINQOGL--BEDDNITERLPENSPIQQENA 569
DB 441 VTNLTKETIKLEKTEDEKAGHKNELNQLSKLNKEKONI-----KNWT 488
QY 570 GLGASPDKPYMIKOVQORYYLAKSIOELIKAK-----DYTKLAKLSN-RHTYNIS-- 621
DB 489 EL-----NDKISLSNEVNLNKDKQTLGNDIKTLNDLNNLNKNEINTSDN 534
QY 622 --LRILKEQLFDVNPRISSRDIEKAFVLDKTEKNKYQIYSSASVPFQNKWSLFGYYRY 679
DB 535 KANKMKEDLAMLN-----EEMEGKCVVIDIYK-KY----- 564
QY 680 LLGLDPKQTIHEL-VKLGQKAGLQEGYENLPSDFNLEDLKN-IRIKTFLFSQKNF-- 734
DB 565 -----KNEIFMLBEKLEK-----ENVADLDELISI--LRNSIYVKEKEFEMKEFYEN 611
QY 735 KLSLDFN-----NYVDGEIKAPFGLPLFIKELRRNSSSGGSSQNSNSPWEQEIISQF 789
DB 612 KINLFNKQFEKKNIYENELNS-----LRLKYDN-----EOGLIKQI 648
QY 790 KDQNLNSQDOLAQFSTKIWEKIIGDENEPDQNNRLQYKLLKDLQESWINKTNDLY-WTY 848
DB 649 DELNIQK-----LK-TEEKYLOLYNDNMFMFRS 675
QY 849 LGDKLVKPKNLEAKFROISNLQELLTAFYTSAAALNNWYQDSGAKSTIIFB----- 904
DB 676 ICTKIDMPYSENIKG-----SOLVDFVTAYIKRDESS-DANPDTHKEMVAELKRRHA 729
QY 905 --IAELDPKVKEKVGADVYQLKPFHYAIGPDDNAGKFNQEV-----RSSSRITLYLKTSK 957
DB 730 AIVAELEEKHKEIA-----KLGEHKEVVLRLGQHKKEETIILSEKHK 773
QY 958 SKLEADTIDQLNAQVKNAPLGLQSFYLDTERFQVFOKLATSLAVQHKQKETL----- 1010
DB 774 ----DVVTKLGEQHKENIILKEECHK-----VVTKLGDQYKEIAKLKESHAVV 819
QY 1011 -----PKLNNNGYTLIHLKLLKXPVTPQIS9SPKQWFB-GKLNQNGSQSNV-----ST 1060
DB 820 VAELEBKHLGEGHKEMVDELEKRAHDFVEGLEEKHAKTAKLEEGHKEEMNEVEKRAH 879
QY 1061 FGSIIESPVSFTNFOEDADLDODGDDSRQNNSLDNQEBAGLLKQKLAILLGNQFIQYYQ 1120
DB 880 FVEGLEEKHKA-----ETAKLG-EGHREVVAGLEEKHKEVVALEEBK-----HKEEIAKLE 929
QY 1121 QNDKETEPIIINVEKVSLSFVFEFK-----LAKTLEDNGKTIIRVLSDETSLIV 1170
DB 930 EGHKEVMAEL--GEKHKEVVAGLEAKHLEEGHKEMVAELKRAHDLAVLEBQHKAEII 987
QY 1171 NTTIETKPEMSAVPEVDFTKWVEQYDPTPLAAKTKFVLKFKDQIPVDGSG-----NISD 1225
DB 988 KLGEHKEHVAGIEEKYKVEATKLAEEHKDQV--TKLGEQHKKEIAKLEDHGHEKVVNEVE 1045
QY 1226 KWLASIPLVTHQ-----QMLRLSPVVTITRELGLKTBQOQOQOQOQOQOQOQKCAVRKEE 1281

Db 1046 KKNASLLNMLBENHKNEMIKL-----KEEHKESASDLVEKLYQKDEE 1087
Qy 1282 LETYNPKDB--FNILNPLTKAHLRLTSLNLV-----NNDPNYKIEDLVKIKNEAGD-HQLA 1333
Db 1088 VANSNNKIEELTNVLKDNLSIMCYKKQILBEEVKRNEYNEINKLIVQNMKMDMDKK 1147
Qy 1334 FSLRANNIKRLMNTPIITFADYNPPFPYNNEDWRSIDKLYLN-----KGNVSSHQ 1382
Db 1148 ILEKENEIKKLNK---KLSNYKVPETKENTYKNSEMVNVENKERIIVDSVCKENISESDV 1204
Qy 1383 QAAGNQSGGLIQRNLKNIKETFTPALIALKDRN-----NTNLNSYSDKIIMIPKYLVE 1438
Db 1205 EGKGN-----LKMTLKK-----KERNIFSINDKNESSELVDTIKSAVI-- 1246
Qy 1439 RSIQVPMSTGLDGYIGSQTQDGTSSSQQKGFQDFTQALGLKNTEVHGKLGISIRFD 1498
Db 1247 -----NKIEMV--KKEIEDNG-----KNIE-----DLNKKILD 1272
Qy 1499 PGNELAKIDASNKKGBE--KLLKSYDLFKNVLNEYKK-----SPKIAKGWTHIHP 1548
Db 1273 LSNELINLENMKVNLTDENNLKKEIEIKDNKLNKEKKNENTEILNLNDDIHK---LKK 1328
Qy 1549 DOKEYPNPKLPENYML-----VLNQPWKTYLN-SSDFITNLVPEPEGSDRGSTKL 1602
Db 1329 EISEWKDEEKLTKENIKLKDIEQINKEYIKKEENLMIKFENI-----NEVTSL 1379
Qy 1603 KOVIQ-----KOVNNY-----ADWGSAYLTFWYDKNIIT-----NOPNVITANIA 1643
Db 1380 KNQIEKMKLEELKNVYELLAEKRETNMISINDDKIVENNILEDTSQNNLNKVE 1439
Qy 1644 DVFIKDVKELEDNKLIAPNITQWMPNIGSGKEFKYPTVPFGNWMENESSMNSQAQPT 1703
Db 1440 DKTGDDINCKNDQA-----KEISYLDKEIKTISMLYGBELNRKNSYDEKVKNLT 1490
Qy 1704 WE-----KIREGPAQALKSSFDQKTRTFVLTTNAPLPLWKYGLFGFQNGPFTQDWRLV 1759
Db 1491 NELKELKIRNKKGBEAI-----ABLNLKL-----NIKEKN-KSV 1523
Qy 1760 FONDNDQIAALRVQODR-PEKSSBDKQKWKIPKVI-----PE--EMFNSGNIRFVG 1811
Db 1524 KQNDSSNNIITKQDGTPEVSNDDKIQDKWKANVLKLEKRPDLWDNINSLEKENFR 1593
Qy 1812 VMQIQGPNLWLPVINSV--IYDFRGYTGDB--SNDVANLVAPWQVKTIAFTNNAFNV 1867
Db 1584 VMSIVKENK--NVQNDKIVGIYSYFKKCEKELKNDMLVICLVKLDILSLFLNDNFVNL 1640
Qy 1868 PKEFN 1872
Db 1641 FEKID 1645

RESULT 6
T28676
rhoptry protein - Plasmodium yoelii (fragment)
C:Species: Plasmodium yoelii
C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004
C:Accession: T28676; A45521
R:Singha, K.A.; Keen, J.K.; Ogun, S.A.; Holder, A.A.
Mol. Biochem. Parasitol. 76, 329-332, 1996
A:Title: Comparison of two members of a multigene family coding for high-molecular mass
A:Reference number: 220507; MUID:97077455; PMID:8920022
A:Accession: T28676
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-2401 <SIN>
A:Cross-references: UNIPROT:Q26216; EMBL:U36927; NID:g1041784; PID:g1041785; PIDN:AAB412
R:Keen, J.; Holder, A.; Playfair, J.; Lockyer, M.; Lewis, A.
Mol. Biochem. Parasitol. 42, 241-246, 1990
A:Title: Identification of the gene for a Plasmodium yoelii rhoptry protein. Multiple co
A:Reference number: A45521; MUID:91101660; PMID:2270106
A:Accession: A45521
A:Status: preliminary

A:Molecule type: DNA
A:Residues: 2260-2401 <KEE>
A:Cross-references: GB:M34281

Query Match 3.8%; Score 366; DB 2; Length 2401;
Best Local Similarity 18.5%; Pred. No. 1.8e-07;
Matches 407; Conservative 341; Mismatches 704; Indels 748; Gaps 106;

Qy 50 IDSVAFKPSIANFTSDYQSVKALLNGKTFDPKSESEFTDFVSKPFDPLTNGRTVLEIPKK 109
Db 430 IDNIKEAKQNYDPQKEHMKTIPTPNEMKYQKPSIEIK--IMKDEFLSKVKN-INDEPKV 486
Qy 110 YQVVISFSEDDOKERF-RLGFHLKEKLDGNTAQSATKFIYLLPL--DHPKAALGQYSY 166
Db 487 YKEKV-----ESBHNKFTLTNKTKEVSDEEIKKENKFNDSKSLNETKKSIEEYQN 541
Qy 167 I-----VDKNFNLIH-----PLSNFSAOSIKPLALTRSSDFI----- 200
Db 542 INTLKKVDDYIKVCLNTNELITCHNKQTTLKOKLNQNIKITKETNSIDKLIYTDKFENIL 601
Qy 201 -----AKLNQFNNODELWVYLEKFPDLEALKANIRLOQTADFSFEKGN----- 242
Db 602 TDKTELETETKTGLSLNNHESNNKELTY--FYD---LKANLGKKNENMLYKQNEKEK 655
Qy 243 -----LVDPPFVYSFTIRNPMQKWEASDLNQDQTVRLYLRTEFSQAK 285
Db 656 AVEDIKKKNVDINKIVSNIETIITYTSYINNEDETE-----NEIGKSIEL-LNTKVLEKVK 709
Qy 286 T-----ILKQYKVKDETFLSSIDLKABNGTSLFANENDLKDOLDVDLLDVSDFG 335
Db 710 ANVTNLNEIKEKLDYDFQD--FGKEKNIK-----YPDENKIKNIDT-----LNQKID 756
Qy 336 GQSETIT-----SNSQVKPVPASERSLK---BRVKPKKQDQ--KPRIEKFSLYBYDALSF 385
Db 757 KSIEITLTKKNSENHIDEIKGQIDKLVKVPKNTMFNEDPKETBKIENTIVEKIDKKKNI 816
Qy 386 YSQQLVELVSPNSIK-DLVNATLARNLRFSLGKY--NFLFDDL-----ASHL-----DY 431
Db 817 YKETDKLLNEISKIENDKTSLEKLNINLSYKSLGNLFLQQIDEEKKABHTIKAMEAY 876
Qy 432 YFLVSKAKIKOSSITTKKLFELPIKISLKSSI-----LGDQE 468
Db 877 IDDLNKKKSQIEKEKNINMDIKMDIHKWKALNISHDDYKLYHTTSKKNHEKISDIR 936
Qy 469 PN-----IKTLFEKVTFKLBFRDV-----ETEKAPGLLYPGVNEELE 507
Db 937 KNSLKIIODFSEESYINDIKKELEKNVLESQNNNTDINQVLSKIENIYNIL-----KLN 990
Qy 508 QAKAQAASFEKSKKGLKEFSQOKEENSKAINNQGLEEDDNITELPENSPIQYQOE 567
Db 991 KIKKI-----IDK-----VKEYTDEIEKNKKKINAE--LSNSEKIITQLKENSLSKECQS 1038
Qy 568 NAGLGASDPKPYM--IKDVONORYLA--KSOIQELIK-AKDYTKLAKLLSNRHTYNIS 621
Db 1039 K--IKSTIDNYSVECKNITNLKTYIVNEKNINITYFKRAEYNO-----NVS 1085
Qy 622 LRLKE-QLFDVNPRISSRDIKAKFVLDKTEKNKYQIYSSASVPFQNKWSLFGYRYL 680
Db 1086 LNFNNIEMADT-----KSQYILNIKQGN----- 1108
Qy 681 LGLDPKOTIHELKVLKAGLOQEGYENLPSDFNLELDKNIRIKTLPFS----- 729
Db 1109 -----GTNN--TDYNIKELKEHKHKKSNYVYKDBAGKNTOEIK 1142
Qy 730 -----QKDNFKLSLLDFNNYDGEIKAPBFGFLPLPKELRRNSSNGSGGSQNSNSWEQ 783
Db 1143 KNKELPKYQGEVTVL-LNKYYAVELK-----NKFDTKNYSE 1179
Qy 784 EIIQSFQKQDN-----LSNQDQLAQFSTKIWEKIIGDENEFQDN-----RL 824
Db 1180 QIIKEIKDAHNFTTSQADKSEKKNMEIKNEQIRIIDEVAKNNKSKAILDQLSVEPFI 1239
Qy 825 QYLLKKLDQ---ESWINKTRDNLWYTLGDKLVKVPKNNLEAKFRQISNLQELLTAFTYS 881

Db 1240 KFLKIDLTAKDCLKTKD-----IETKISNLSIDTQETKLTENKILNTLEKLES 1293
QY 882 AALSNNVYQDSGAKSTIIFEBIABDKVKEKVGADYQYQKLFHYAGF----- 931
Db 1294 --LKNQKKNIEQK-----KELDEVNSKIK-NIESNVQHKQNGVEIGVEKINEIAKA 1343
QY 932 -----DNVAGKFN-----QEVTRSSR-TI 950
Db 1344 NKQIBESTOKLIPTIKNLISPPKANDLEGIDTNKMLGKYNTEMNNIYEEFIKSYDLITH 1403
QY 951 YLKTSCSKLEADTI-----DOLNAQVKAPGLQSFYLDTERFGVQFKLATSLAVQ 1002
Db 1404 YLETVSKEPIITYEQIKKRIITAQNELLTNIK--VNKAQSYLDDIEANFDRIVH---- 1457
QY 1003 HKQKERTLPKLN--NDGYTLIHDKLKP-----IPQISSPEKDFEGKLNQNGQSO 1054
Db 1458 -----FRKLNVDNDKFTNEYSKVNKGFNINSINNKKSTDENLLNLINQTK-- 1508
QY 1055 NVNVSTFGSIIESPYSTNPFQEDADLDQDQDSDROCN-----NSLDNQEAGLLKQ 1105
Db 1509 -----MYANIVSKYYSYKYEAE-----NIFINIPKLANGSLNIQ-----IKS 1545
QY 1106 KLAI-LLGN--QFIOYQONDKIEFIEIINVEKUSELSPRVEFKLAKTLEDNGKTIRVL 1161
Db 1546 SSGIDLFKNINIAILPYLSQKQKDTLTFIPSPKETSSETVTKI-----SDSYNTLLDI 1597
QY 1162 SDTMSLIVNTTIEKTPENSAVEPDTKWVEQYDPRTPLAAKTFVLKFKQDIPVDGSG 1221
Db 1598 LKESQEL-----QKKEQALNLIIFENLLHD-----KVOATNELKDTL--SDLK 1639
QY 1222 NISDKWLASIPVHQ--QMLRLS-----PVVKT-----IRELGKTEQOQOQ----- 1262
Db 1640 NKKEQILNKVLLHLKSNELNKLSCSNQNYDTILES KYDKIKESNNYKEKENIGINF 1699
QY 1263 --QOQOQOQOQKAVKBELETYPKDEF-----NINPLTKAHLTILSNLVNNDP 1313
Db 1700 DVKAMEEQFNQDKIEKLENNYKHEKDNYPSENNILQSKKIKELT--NAPNAEI 1757
QY 1314 NYKIEDLVKVN-----BAGDQHLAFSLRANNIKRLMNT--PITFADYNPFF-----YV 1360
Db 1758 K-KIEDKIEKGLINKLIEKDCMLFT-----YKTLVETLKIKTDTYTKFITSATKFS 1811
QY 1361 NEWRSIDKYNKNGVSSHQQAAGNQSGGLIQRINKNIKPETTPPALIAKDRNNTN 1420
Db 1812 KBFKLYDATSNLSNDDINTLOTKYDLNQ-----INKHV-----ASVADATNDNNN 1858
QY 1421 LSNYSDKIIMIKPKYLVERSIGVPWSTGLGYIGSEOTKDGTSQQKGFQDQDFOALG 1480
Db 1859 -----LIEKE-----KEATKT-INNLTELFTIDSNKIDADG 1888
QY 1481 LKNTVEHGLGLSIRIPDPGNELAK-----TKDASNKKGEKLL-----KSYDLFPNY-- 1528
Db 1889 LHNKK-----IQIYFNSLHKSIDSIKQYKQWHPKFLNIGHINKYFDISKEFD 1940
QY 1529 -----LNEYEKSPKIAGWTHIHDPQKEYPNPN-QKLPENY----- 1564
Db 1941 NILQOSELNTANLNDLSEIKQKISDKKQFLHALSETPIPNFTLKEIYHDIVKYKNGQI 2000
QY 1565 --LNLVLNQPWKVTLYNSDFTINLVEPEGSDRGSTKLVKOVIOKVNNYADVGSAY 1621
Db 2001 DEIENITNEENITLY--MDII-----TKMKVQSILN-----F 2034
QY 1622 LTFWYDKNIITNPQNVITAMIAVFIKDVKELEDNTKLIAPNITQWMPNITSGSEKPYKP 1681
Db 2035 VTTY-----ENDSNIIKQHIQDNNENDVSKIESLETTIQSFQKILNKLNEIKAPFY-- 2086
QY 1682 TVFFGNENENSSNNSQAQPTW-----EKTRREGFALQALKSSFDQKTRFVL 1730
Db 2087 -----DNNNINNVISTISQDNDVKKHISKDLTIENELTIQIKSLEDIRKSTYDIR 2137
QY 1731 TNAPLPLWKY-GPL-----GFQNGPNFKTDQWRLVFQNDNDNQIAALRVQEQDRPEK 1780

Db 2138 SS---QITKYVNPIDHYVEQQTKKIQNNPN---KD-----EIDD-----LIQEIVNTNK 2180
QY 1781 SSEDKDKQKWKFKVVPPEENFNSGNIRFVGVMQIQGNTLWLVPVINS-SVIYDFYRG TG 1839
Db 2181 ESELK-----LPTIINNKNV-----TPIISRIDKVINLIKSEY 2214
QY 1840 DSDNVANLVAPWQVKTIAFTNNAFNNVFKBFNISKKIVE 1879
Db 2215 NNNDVSVTVAK--KLEEDANNIIRDLDTSNNMLNDLIQ 2251

RESULT 7
Ti8501
hypothetical protein C0760c - malaria parasite (Plasmodium falciparum)
C:Species: Plasmodium falciparum
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004
C:Accession: Ti8501
R:Lawson, D.; Bowman, S.; Barrell, B.
submitted to the EMBL Data Library, August 1997
A:Reference number: Z18935
A:Accession: Ti8501
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-3394 <LAW>
A:Cross-references: UNIPROT:O77384; EMBL:Z98551; NID:e1331903; PID:e1331922; PIDN:CAB1114
C:Genetics:
A:Map position: 3
A:Note: C0760c

Query Match 3.5%; Score 340.5; DB 2; Length 3394;
Best Local Similarity 18.3%; Pred. No. 3.5e-06;
Matches 390; Conservative 307; Mismatches 791; Indels 641; Gaps 87;

QY 59 IANFTSDVQSVKALLNGKTPDKPSSEBTDPSKDFDLTNNGRTVLEIPKPYQVVISERS 118
Db 1164 IHNIEDKYKVVYFENLFNFN-----FSNVLSFDLLIRRPDKILRLTKYEQELL---- 1214
QY 119 PEDDKERFRLGPHLKEKLEDGNIAGSATKFIYLLPLDMPKAALGOVSYIVDKNFNNLIH 178
Db 1215 -BENYEK-----IKNKEEYELHACIK-----ELEMNLERYNNEKIILDSQINE--- 1258
QY 179 PLSNFSAOSIKPLALTRSSDFIAKLNQFNQNDLWVYLEKFFDLBALKAN--IRIQADP 236
Db 1259 -----KEKKNINIENEKYLILEKEYE-EYQNKNIPIFNAQIENL 1294
QY 237 SPEKGNLVDPPVY-----SFRPNQOKEWASDLNQDQKTVRLYLRTFESFQ 283
Db 1295 EKEKKQLOBEIIQKOMINVKLNKKNCDIIKIYEKEQYLLTLQENKDSHNYLKOKFE-N 1353
QY 284 AKTILKQVKYKDETFLSSITDLKASNGTSLFANENDLQDQLDVLDDVSDYFGQSETITS 343
Db 1354 LNLNLEKLYDHDISLUNKI-----NTLWLEHKKNNKNTFHMNNLEVEN-----N 1398
QY 344 NSQVFPVPASERSLKDVRKFKDQOKPRIEKPSLYEYDALS----- 384
Db 1399 NLLLK-----MKELQKNYNIKKELNERIKQINFRNNVSTLSLRNDRSTRGSIHOINNM 1453
QY 385 -----FYSQLOELVKPNSIKDLV---NATLAR-NLRESLKG 417
Db 1454 YMNTHLFGMGASKINNINSLYNSNMTHMRGSIIRKNKEBAEGNSTQARNNKDSTON 1513
QY 418 Y-----NPLFDLASHLDYVFF-----LVSKAKIK-QSSITKK 448
Db 1514 IINNHTONINNMNNINNTLNSINSHLYYFPFPHNNVNSPKVMGCMCDVTLASGVNKK 1573
QY 449 -----LFTIEPIKISLSSILGDQEPNI-KTLFPEKVTFKLONFRDVEIEKA 494
Db 1574 DQFLNLEENEENSFLEYIRIKSLQEELCDKSEILKIKGKNILITC-----IETWKC 1628
QY 495 FGLLYPGVNEELEQAKAQRASFEKSKGLKEF---SQOKEENSKAINNQEGLED-- 549
Db 1629 F---CKNSKEETISRLKEICEQLEKH-----KEFLINKNSNEDKUKYINSLSCDEKDY 1679

1368 LLGQTINALVTPNGEKIFSDTELQSLTSLINESLDFGKTNIIISNDYDLVLSALSIQES 1427

1314 NYKIEDLVTKNAGDHQALAF-----SLRANN-----IKRLMNPITFA--D 1353

1428 SSVLTKEVNTTHKALDOONSYPFAVITKLSVKEGDVDSDEPKAIIKKELN---LFAKND 1484

1354 Y-----NPPFYNYEDWRSIDKYLNNKGNVSSHQQOAGAGNQGSLIQBLNKNKIPETPT 1407

1485 YLLTKFTSPI-----VDKTIIDFDSATKSS-----IVKLIKNV----- 1516

1408 PALIALKORNTNLS-NYSKDIIMIKPKYILVERSIGVPMWSTGLDGVIGSEQTKDGTSSSS 1466

1517 -----LSDEQNLFKSAQIIDTIFANKKKIYAYENLN---DFLIDALVDNKNINILNUSNT 1568

1467 QOK-GFDQDF-----IOALGLKN-----TEYHGKLGLSIRIPDPGNELAKIKDASNKKG 1514

1569 LGKIKTDDEFKNIKSPITKNLMTSDSSITEBIN-----TIVKSAHELISIVDSSNFLN 1623

1515 EEKLLKSY-----DLFKYLYNEYKKSPKAKGWT 1544

1624 Q--LLKDFALLEKNKKDKSTSLFSTFMKKSISBFLDAENNYLALFKK--VLSSNLL 1678

1545 NIHPDQKEYPNQKL-----PENLYNLVNQPK-----VTLYNSSDFITNLFVEPE 1592

1679 STOSDQSSFTKSISSILRVFLRKEKALEMLKKOFRNFELHDIKVDADALNLIKPIFRDQ 1738

1593 GSDRSGTKLKQVIOKQVN--NNYADGSAYLTFWYDKNIIITQNPVITANTADVIKDV 1650

1739 VID-----FIETIINRVVSQKEYSKLNS-YQEILY-KFLSSKNKETL-----NPFKKQI 1786

1651 KELEDNTKLIAPNITQWPN-IGSKEKFKYKPTV-PFGNWNEN-----1692

1787 GEIK-NSKLINPLAKSPISIELKSKITLQNDIIDSIVNFVNESLVLEKTDLIDKLYDL 1845

1693 --SSMNSQAOPTWEKIREGFALQAKSSFDQKTRTFVLTITNAPLPLWKYVGLGFONGFN 1750

1846 WVSFNAKVSDKTKVEIFNFSSL--LKPIMQDKNTYFSLVKTLATINWKLI--MESQNTDA 1901

1751 FKT---QDWRLVDFONDI--NQI---AALRVQEOQRPEKSESDKDKQWKIKFKVVIPEEMF 1802

1902 LKNALQGMILRVFKNTDLLNKFDPVLQNFIDIDQDKSKTKN-----ENLR 1949

1803 NSGNIRFVGVMQIQGNPTLMLPVINSVIVD-FYRGTGDSNDV-----ANLVAPWQVK 1855

1950 NASQOILVFVLQNEBDTEFFKSLVNNIIISNNQYKANNYYLDIVSFIEKNKLVQYFK 2009

1856 TIA-----FTNNAFNVPKEFNI 1873

2010 KIAPKIASDSMLKTFVNGFALNLLQEQNL 2038

RESULT 12

hypochemical protein MYPV 4140 [imported] - Mycoplasma pulmonis (strain UAB CTIP)

C:Species: Mycoplasma pulmonis

C:Date: 24-May-2001 #sequence_revision 24-May-2001 #text_change 09-Jul-2004

C:Accession: F90563

R:Chambaud, I.; Heilig, R.; Ferris, S.; Barbe, V.; Samson, D.; Galissou, F.; Moszer, I.; Nucleic Acids Res. 29, 2145-2153, 2001

A:Title: The complete genome sequence of the murine respiratory pathogen Mycoplasma pulmonis

A:Reference number: A99512; MUID:21267165; PMID:11353084

A:Accession: F90563

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-2244 <KUR>

A:Cross-references: UNIPROT:Q90F02; GB:AL445566; PID:gl4089828; PIDN:CACI3587.1; GSPDB:Q100000000

A:Experimental source: strain UAB CTIP

C:Genetics:

A:Gene: MYPV 4140

A:Genetic code: SGC3

Query Match 3.2%; Score 312; DB 2; Length 2244;

Best Local Similarity 18.7%; Pred. No. 3.1e-05;

Query Match 3.2%; Score 312; DB 2; Length 2244;
Best Local Similarity 18.7%; Pred. No. 3.1e-05;

Qy 913 KEKVG-----DYYQLKPHYAIGFDDNAGKFNQEVIRSSRTIYLKTSQ-KSKLEAD-TID 966
Db 909 NEKNEIISVFDIY-----SDIFGLIDEII-SNNKYLANEIDGFYVKEIDPTTG 957
Qy 967 QLANQAVKAPLQSFYLDTERFGVQKLAISL-----AVOH 1003
Db 958 LLSVLKK---GVYRGFYPTDRIGIIPLLKASAKFKTTGINYLYKVGTHBYGHHTLOQY 1014
Qy 1004 KQ-----KEKTLPLKLNNDGYTL--IHDKLLKFPVPOISSPEK-----DWFEKLNON 1050
Db 1015 AQDVSDPSVPIPTAIGRNGISSHLHNDVLQVLDARSSGLKVRKKNIPY---KPNR 1071
Qy 1051 GOSQNVNSTFGSIIESPYFSTNFQEDADLDQGDQD-----SRQGN 1092
Db 1072 GVLPNYTKLLDLSLKS-----QSGKNDHGSHSNEDIWEKIDIFGSENN 1117
Qy 1093 NSLD-----NOEAGLLK-----QKLAILLGNOP-----IQ 1117
Db 1118 KTLDSNDKVLSENERFLQTVVVGIEKAAKARNLQYDLFILNSFDHESATINPGISGKE 1177
Qy 1118 YYQONDKEIERPIINVKVSELSFVFEKLAKTLEDN-GKTRV-----LS 1162
Db 1178 YFKLDK---VEFFPAEDENSTDIKSFPEKTDVLSNGLNVKPKDKRKIIIAITYKEDLA 1234
Qy 1163 DETMS-----LIVNTTIEKTPMSAVPEVDFDKWVEQYDPTPLAAKTFV- 1208
Db 1235 NDTFSDLKPTTPIAKDKPLDLSIFEKTKSLKALKKEIEK-----QRLLENEPEN 1285
Qy 1209 LFKPQOIPVDG-SGNISDKWLASIPLVIHQWMLRLSPVVKTRRELGLKTEQOQQOQQOQ 1267
Db 1286 LDFQTTFSVNGWSLKKDLTKIB-VIHEE-----NNELNENHLODFK 1328
Qy 1268 QOQPOKKAVRKE-----BELETYNPKDEFNINLPLTKAHLRTLSNLVNDN-N 1314
Db 1329 EFVETGFKIKKYVDFTKLENFYQYLRKHGKANFTF-----YQLLSTLITGQVKT 1381
Qy 1315 YKIEDLVKVNKAGDHQALFSL-----RANNIKRLMNTPTT-FADYNPFYFYNED----- 1363
Db 1382 LKGKEYLQNDQDELWLRAKILHDPSRFNGIDELIKQRTSYYPNYSIFHPENADYFN 1441
Qy 1364 -----WRSIDKVLANKGVSSHQOQAGNGOGLIQRLNK----- 1399
Db 1442 PRKENRLKQOIRGKKNYSLDLALSNYKONSHENE-----NKKATHY 1487
Qy 1400 -----NIKPETFTPALIALKDRNNTNLSYSDKIIMIKPKYLVERSIGVPMSTGLDG 1451
Db 1488 YGGEFNEFKVKIDGKEPIILDWDYEHFNFGDNHFKENFFRTK--ENESI-----TSFDS 1540
Qy 1452 YIGSBQ-----TKDGTSSSSQO-----KGFDDQDFIOALGLKNTE 1485
Db 1541 FSPFPFSSIDPLKFTLTGQGENAQRIMDIDYALTKFNLFDYAKKLDSDLFKTTLDQSK 1600
Qy 1486 YHGKLGLSIRIPDPGNELAKIKDANKKGEELKLSYDLFKNYLNEYEKSPKI----- 1539
Db 1601 KYDKAAL-IOAPE-----KAYGOEKAKEL---QDIANDLMDKFESSLAFPMKNHN 1648
Qy 1540 -----AKGWTNTHPQKEYPNPQKLLENYLNLVNLQWPKVTLVNSSDFI 1584
Db 1649 IKDLSVDHLFVSOVILGFNNITKSLR--PNPNDN-----NWVIER-WKTSPISSDVI 1699
Qy 1585 TNLFVEPEGSDRGSTKLGKQVTKQVNNNYADWGSAYLTFWYDKNIIITNQPNVITA-NIA 1643
Db 1700 LRAKTDNPSND---SRIVMDAIRKFLKDNNDI--DSKLDLFLKYLILGTQ---ILNGKND 1752
Qy 1644 DYFIKD---VKELED-NTKLIAPNITOWP-----NISGSKFKYKPTVF----- 1684
Db 1753 QIFIPNNFOINELNKNKISRFENDFGQYFSDVFNFAESLRTDYVQVTSYVPSQKQLKNN 1812
Qy 1685 -----FGWENENSNMNSOQPTTWKIREGFALQALKSSFDQKTRTFVLTNAPLPLWK 1739
Db 1813 GYLKGFSEWNTGNEYVSGSLTKW-----LNQINVLSD-----LLRRRSIYSRWS 1859

Qy 1740 YGP-----LGFQGNPFKTDWRLVLFQNDNDQTAALRVQE-----Q 1775
Db 1860 YVEQASQFLHFGNAQVFNYSKKL-----DEVLSTLTKEFKFDQSQSPLVNYFELDSYQ 1914
Qy 1776 DRPEKSESDKQKWKIKVVIPEEMSPNSGNIRFV 1810
Db 1915 EKQKLNSESTKK-----TEELDRSDLKYYI 1940
RESULT 13
S38173
myosin-like protein MLP1 - yeast (Saccharomyces cerevisiae)
N:Alternate names: protein YKR095w; protein YKR415
C:Species: Saccharomyces cerevisiae
C:Date: 03-May-1994 #sequence revision 03-May-1994 #text_change 09-Jul-2004
C:Accession: S38173; S40647; S31207
R:Baladron, V.; Ballesta, J.P.G.; Bou, G.; del Rey, F.; Eteban, P.F.; Garcia-Cantalejo, submitted to the Protein Sequence Database, March 1994
A:Reference number: S38158
A:Accession: S38173
A:Molecule type: DNA
A:Residues: 1-1875 <BAL>
A:Cross-references: UNIPROT:Q02455; EMBL:Z28320; NID:9486586; PID:9486587; MIPS:YKR095w
A:Experimental source: strain S288C
R:Bou, G.; Eteban, P.F.; Baladron, V.; Gonzalez, G.A.; Cantalejo, J.G.; Remacha, M.; Jin Yeast 9, 1349-1354, 1993
A:Title: The complete sequence of a 15 820 bp segment of Saccharomyces cerevisiae chromosom
A:Reference number: S40644; MUID:94205265; PMID:8154186
A:Accession: S40647
A:Molecule type: DNA
A:Residues: 1-1875 <BOU>
A:Cross-references: EMBL:X73541; NID:9450550; PIDN:CAA51948.1; PID:9450554
A:Experimental source: strain S288C
R:Koelling, R.; Nguyen, T.; Chen, E.Y.; Botstein, D. Mol. Gen. Genet. 237, 359-369, 1993
A:Title: A new yeast gene with a myosin-like heptad repeat structure.
A:Reference number: S31207; MUID:93247549; PMID:8483450
A:Accession: S31207
A:Molecule type: DNA
A:Residues: 1-300, 'A', 302-1875 <KOE>
A:Cross-references: EMBL:L01992; NID:9171958; PIDN:AAA34783.1; PID:9171959
C:Genetics:
A:Gene: SGD:MLP1
A:Cross-references: SGD:S0001803; MIPS:YKR095w
A:Map position: 11r
Query Match 3.1%; Score 305.5; DB 2; Length 1875;
Best Local Similarity 19.6%; Pred. No. 4.5e-05;
Matches 340; Conservative 271; Mismatches 627; Indels 494; Gaps 75;
Qy 60 ANFTSDYQSVKALLNGK-----TF-----DPKSSEFTDFVSKFDFL-----TNNGR 101
Db 380 AKSSDDIFLKLQLIKERRTKHEHLQNIQTFIVELEHKVPIINSFKERTDMLNELNAA 439
Qy 102 TVLEIPKQYQVIVSEFSPEDDKERFLGRFLHKEKLEDNIAQASATKIYLLPLDMPKAA 161
Db 440 LLE-----HTSEKNNAKVKELNAKNQKVECNLDQLTK-----QRLDLCR--- 482
Qy 162 GOYSYIVDKNFNNLIHPLGNFSAQSIKPLALTRSSDFIAKLNO-----FNNQDELWYVL 216
Db 483 -QIQYL-----LITNSVNSDKGLR---KEEIQFIQNIQEDDSTITESDSQKVV 530
Qy 217 EKFPDLBALKANRLQADFSFEK--GNLVDPFVYSFIRNPQKQEWASDLNQDKTVRL 274
Db 531 ERLVEFKNI---IQLEKNAELLKVVNRNLADKLEKSKQSLQKIESETVNEAKEAI 587
Qy 275 YLRTEFSPQAKTILKDYKYDE--TFLSSIDLKASNT--SLPANENDLKQDQVLLDV 330
Db 588 TLKSEKVDLSRIEELQLELEKTSVPNEDASVNTVIKQLTETKDDLESQVQDLQTRI 647
Qy 331 SDYFGGQSEIT--SNSOVKVPASERSLKDRVFKQDQKQPRIKFSLEYDALSFYSQL 389
Db 648 SQITRESTENSLNKIEQDLYDSKSDISIKLGKESRILAEERFKL----- 695

Qy	320	-----KQOLOVDLID-----VSDYFGGQSETIT--SNSQVFPVPASERS	356
Db	1059	LESVIAEKEQLTKDLENIENTIENOEELRLGLDELKQOEIVAOEKNAHAIKEGELSRT	1118
Qy	357	LKDRVFKDQOQPRLEKFSLYEVDALSFYSOLOELVSPNSIKDLVATLARNLAFSLG	416
Db	1119	C-DRLAEVBKLEKSGOLOEQOQLNVQEEMSEMQKINELNKN--LKNKELTLE	1175
Qy	417	KYNPLFDDLASHLDYFLVSKAKIKOSSITPKLFIELPIKISLSSILGDQEPNIKTLE	476
Db	1176	HMETERLELAQKLNNYEEVKSITKERVKLKEL-----QKSPF	1213
Qy	477	KEVTFKLDNPRDVEIEKAFGLIPGVNHEELEQAKAQRASFEKEKSKGLKFSQOKEEN	536
Db	1214	TERDHLRGVIREI-----ATGL-----QTKHEELKIAH-----IHUKHEQETIDELRRSVSEK	1261
Qy	537	SKAINNOEGLEEDDNITERLPENSPIOYOENAGLGASDPKPYMKDVONORYLAKSOI	596
Db	1262	TAQIINTQDLEKSHT---KLOEEIPVLHBEQEL-----LPNVKVSETOETMNELELL	1311
Qy	597	QELIKADYTKLAKLJASNRYNISRLKEQLFDVNPRIPS--SRDIEKAFVLDKTEKNK	655
Db	1312	TEQSTTKDSTLARIEMER-----LRLNEKFPQESQEEIKSLTKERDNLKTIKEALEV-K	1364
Qy	656	YWOIYSSASPVFNQKWSLFGYYRYVLGLDPKQTHIE--LVKLGQKAGLOFEGYENLPSDFN	714
Db	1365	HDQL-----KEHIRETLAKIOESQSKQ-----BQSLN	1391
Qy	715	LEDLKNIRIKTLPFSQKDNFKSLLDFNNVYDGEIKAPFGLPLFELPKELRRNSSNGGS	774
Db	1392	MKEKDN--ETTKIVISEMEQPKPK-----DSALLRIEM--LGSLKRLQESH--	1435
Qy	775	QNSNSPWEQETISOFQDNLSNODLAQAFSTKIWE---KIIGDENEPDQNNRLQYKLLK	831
Db	1436	-----EMKSAKEKDDQLQELVQESDQLKENIKEIVAKHLETEEBELKVAHCCLKE	1488
Qy	832	LOESWINKTRDNL-----WYTIQDKLVKPKNNLEAK-----PROJISNLOELL	875
Db	1489	-QEBTINELKVNLSKEKETESTIQKQLEAINDLQNKIOEIVYKEEQOLNKIOISEQENV	1547
Qy	876	TAF-----YTSALSNWNNVYODSGAKSTIIFREIAELDPKVEKVCADV	920
Db	1548	NELQKFKEHRKAKDSALQSTESKMLEITNLRLQESQEEIQIMIKEEEM--KRVQEAQIER	1606
Qy	921	YOLAFHYAIGPDDNAGFNOEIVRSSRTI--YLKTSQKSKLEAD--TIDOLNOAVKNAPL	977
Db	1607	DOLK-----ENTKEIVAKKESQEKVQFLKNTAVNETQEKMCIEHLKQGFQTKL	1658
Qy	978	GLQSPYLDTERFGVQFKLATSAYOH-----KQKEKTLPKKLNNDGYTLIHDKLKK	1028
Db	1659	NLEN--IETENIRLTQILHENLEEMSVTKERDOLRSVEETL--KVERD---QLKENLRE	1711
Qy	1029	PVIQIOWSSPKDWFEGKLNQOSQONNVSTFGSIIIESPYFSNFOEDADLDODG----	1084
Db	1712	TITRDLEKQEBELKIVHMHLEKHEH---QETIDKLRIVSEKTNBISNMQKLEHSDNALKAQ	1768
Qy	1085	---QDSSRQGNNSILDQ-----AGLLKQKLAIL-----GNQFIQYQONDKEI	1126
Db	1769	DLKIQEEURLIAHMLKEQOETIDKLRGVSEKTKLSNMQKDLNSNAKLOEKIQLKAN	1828
Qy	1127	EFEIINV-----EKVSLSFRVFPKLAKTLEDNGKTIIRVLSDETMSLIVNTTIEKTP	1178
Db	1829	EHQILITLKDVNETOKKVSEME---QLKKQIKQDSILTSLKLEIENLNL--AQELHENLE	1882
Qy	1179	EMSAPVEPVDT--KWVEQ-----YDPTPLAATKTVLKFQDOIIVDG	1219
Db	1883	EMKSMKERNLRRVEETLKLERDQKESLOETKARDLEIQOELUKTARMLSKHEKHTVD--	1941
Qy	1220	SGNISDKWLASIPLVHQMLRLSPVVKTIIRELGKTEQOQOQOQOQOQOQOPKAVRKE	1279
Db	1942	--KLURE-----ISEKTIQSDIQK-----DLDESKBELQKIKIQLQKELQOILLRVK	1986
Qy	1280	BELETYNPKDSFNILNPITKAHRLITLSNLVNDPNYKIEDLVKINBAGDHQLAFSLRAN	1339

RESULT 16

RES01
S05603

S05603
major merozoite surface antigen precursor - malaria parasite (Plasmodium falciparum) (str
N:Alternate names: gp195 surface antigen
C:Species: plasmodium falciparum
C:Date: 12-Feb-1993 #sequence_revision 12-Feb-1993 #text_change 09-Jul-2004
C:Accession: S05603; S04850
R:Myler, P.J.
submitted to the EMBL Data Library, April 1989
A:Reference number: S05603
A:Accession: S05603
A:Molecule type: mRNA
A:Residues: 1-1639 <MVL>
A:Cross-references: UNIPROT:P04933; EMBL:X15063; NID:g9896; PIDN:CAA33163.1.; PID:g9897
R:Myler, P.J.

Query Match	3.0%;	Score	235;	DB 2;	Length	1639;			
Best Local Similarity	19.5%;	Pred. No.	0.0001;						
Matches	282;	Conservative	204;	Mismatches	478;	Indels	484;	Gaps	68;

QY	203	LNQFNQDELWVYLKFPDLEALKANIRLOTDAPFEKGNLVDPFVYSTRIPQONKEWA	2626
DB	168	IDGYEINELLYKLNFPYDLLFAKLN-----DVCANDYQCQIPFNLIKIRA	2111
QY	263	SDLNQDQKTVRLYLRTETSPQAKTILKDYKYDETFLSLDLKASNGTSLFANENDLKQ	3222
DB	212	NELDWLKLVL-----FGYRKPL-----DNIKDN	234
QY	323	LDVDLLDVSDFGGOSETITNSQVKPVPASRSILKDRVFKKDKQOKPRIEK-----FSLY	3787

Qy	482	----	KLDNFRDVBIEKAFGLLIPGVNNEELEQAKQAQASFEKEKSCK	524
Db	480	SRTQCSDLGREVICLMAEHLHNETKSRNVATPVQVALDEYAQNPTASSETLVNKE----	535	
Qy	525	GLKEPSQOKEBNSKALNNOE--GLEEDDNITERLPENSPIOVQOBNAGLGAPOKPYMI	581	
Db	536	-LANPSSIIKEAVSKTLELREKVPALBCDVEI-----OKQTVQYQISNAKENSNTLSEOI	589	
Qy	582	KDVQORVYLAQSI-----QELIKADVTKLAKLIS-NRHTYVNIISRLIKEOQLFDVN	632	
Db	590	KNLESE--LNSSKIKNESILNERNLLKEMLATSRSSILSHNSAGNIDDKMKS-ID--	642	
Qy	633	PRIPSSRDIEKAFVLDKTEKNKYWOIYASASPVFQNKSLP-----GYRVLLGL	683	
Db	643	--ESTRELEKNYEVY-----RNMETAQIESLKRNDLISEMEAIKELENSKYOQOOLST	695	
Qy	684	DPKQTTHELIVKLQKAGLOFEGYENLPSDFNLEDLKNIRIKTPLFSQKD-----NFKLSLL	739	
Db	696	DRUTNANNDEAFKK-----EAKELRSINQLOD-----IISRQORASKFAEELL	741	
Qy	740	DFNNYD--GEIKAPEFGFLPLPKELRRNSSGSGSONSPSPWBOE-IISQFKD-QNL	794	
Db	742	HVNSLAERLKGELNASKG-----EXDLKRTQERLISENDKLLAERERLMSLVSDLOTF	795	
Qy	795	SNODQLAQFSTKIWEKIIIGDENEF-----DONRLO--YKLLKDLOESWINTRON	843	
Db	796	LNQQLSDAARKV--KPESEKESLSLSLOKLKESNEKMSNDLHSLQKLEKSIGEYS--S	851	
Qy	844	LYTTYIGDKLVKVPKN-----LEAKPROISNLQEL-----LTAFYTSAAALSNKW	888	
Db	852	RIKTLMLEKOSTSEDNRKLLDNQOMMEIKLOELNGVIELEKQBFSTLEAKFTQ--OKNT	908	
Qy	889	NYQDSGAKSTIIFEBIAELDPKVEKVGADVQLKPFHVAIGPDDN--AGKFNQOEVRS	945	
Db	909	SYSE-----REALLESSLDLOSHTS-----LESQYNSLRNIEQLOAASKLAEWV-E	957	
Qy	946	SSRTYV---LKTS---GKSKLEADTIDQ--LNOAVKNAPLGLOFSYLDTERFGVQK	994	
Db	958	RVKTEYDEYRLQTSLESKONHLKITSLQERIVILQDEBIASSLRCEINITHDSE-----	1010	
Qy	995	LATSLAV-----QHKQEKETLPKLLNNDGYTLIHDKLAKKVPQI	1034	
Db	1011	--TRVALLLEENKHLNNELSHRNAEKQHLKENDYKQOL--LLVTEDLRK-----	1057	
Qy	1035	SSSPEKDFEGLKNGONGSONVNVTPGSIIESPYFTNPFQEDADLDQD-----GODDSR	1089	
Db	1058	---TRDYEKELLRHADAR---STLOKURED--YTKALEQVEDLNKEITALKAGINESQ	1107	
Qy	1090	-----QGNNSLDNOEAGLLKQKLAAILGNOFIYIYOQNDKEI-----	1126	
Db	1108	PPPISEKEDPL-RQEVVYVLKQNAWLLTQ--LQSSNLNFAEITSPSPDLDSVMKGLSLD	1164	
Qy	1127	-----EPPIINVBKVSLSPRVEFKLAKTLEDNGKTRIVLSDETMSLIVNTTIEKT	1177	
Db	1165	QNHVKRISKEMEIISCOR--QLLPLENKLRVTVESSNRVIADLQRGITEKDVSSSTESV	1222	
Qy	1178	PEMSAVEVPDTKWVEYDPRTPLAATKFKVLKFKDOI PVDGSGNISDKWLASIPLVIHQ	1237	
Db	1223	GRS-----NYLNNVALLNESNKS-----LRE	1244	
Qy	1238	QMLRLSPVVKTIREL--GLKTEQ--QOQOQOQOQOQOQKQAVRK-----EBELETYNPK	1288	
Db	1245	NLERNEEVITELREKIETKTDLANFRNLKEQLESQLOTEKAAVKKLENSNEEYKRNQE	1304	
Qy	1289	-----DEFNILNPLTKAHRITLSNLVNNDPNVKIEDLKVKIKNAGDHQLAFS	1335	
Db	1305	ILLSNGSSTSSDASRLKNELVSKENLI--EELNQBIGHLKSELEYTKSKSEDLN--N	1359	
Qy	1336	LRANNIKR-----LWNTPIITEADYNPPFYNNEDWRSIDKYLNNKGVNSSHOOQAAGNQ	1390	
Db	1360	ERAQNOQSKIEOLEKNTKLA---WRTKYEVQVNV-SLEKNNQIQOOLSQK	1407	
Qy	1391	SGLIO-----RLNKQIKPETFPTPALIALKDRNNTNLSNYSDKIIMIKPKYLVERSIGV	1443	

Db	1408	TSELEAKVAECHQLNEQLNKPSTATTATQSEPSVTSLEEFN-----	1449
Qy	1444	PWSTGLDGYIGSEOTKOGTSSSSQOQGFDDFIQALGLKNTYHGKLGLSIRIFDPCNEL	1503
Db	1450	-----STKEELSSTORK-----LSEIMDILNT-----TKDEL	1476
Qy	1504	AKIKDASNK-----GEEKLLKSYDLFKNYL-----NEYEKKSPKIAK	1541
Db	1477	EKVQNSNKSEGTSKDTEIPNEEMERKVMQOEVLRBSRIAKELQKELLARKQNOVLQ	1536
Qy	1542	GW-----TNIHPOQKEYPNPNQKLPENYLNVLNPNQKVTLYNS--SDF	1583
Db	1537	DQVKALQETVVSSEBASVHADTKOL--BNLKKTTEMLSVTFQ-----VIFNESISDF	1589
Qy	1584	ITN-----LFVEPGSDRGSTGLKQVITQKQVNNYVADGWSAYLTFWYDKNIITNQPNVI	1638
Db	1590	STSTADFTTFVQKEWEK-----REILQKVVEEQVAOSHQKOLD-----	1628
Qy	1639	TANTADVPIDKVELEDNTKL-----IAPNITQWPNISGSKKFKYKPTVFFGNWENENSS	1694
Db	1629	--NIR-----KELEMNKULSLMLEKMLARVRAELESQKK--DSPAILLSLEASKNTDS	1678
Qy	1695	MNSQAQTPWEKIREGFALQALKSSFDQKTRTFVLTNAP	1734
Db	1679	NKSNSVP-----AAQVKEKKLIAKTHS--VDTNSP	1707

RESULT 18
 A24594
 Probable major surface antigen (83K, 19K, 42K) precursor - malaria parasite (Plasmodium falciparum)
 C:Species: Plasmodium falciparum
 C:Date: 29-Aug-1987 #sequence_revision 29-Aug-1987 #text_change 09-Jul-2004
 C:Accession: A24594
 R:Holder, A.A.; Lockyer, M.J.; Odink, K.G.; Sandhu, J.S.; Riveros-Moreno, V.; Nicholls, S.
 Nature 317, 270-273, 1985
 A:Title: Primary structure of the precursor to the three major surface antigens of Plasmodium falciparum
 A:Reference number: A24594; MUID:86014355; PMID:2995820
 A:Accession: A24594
 A:Molecule type: DNA
 A:Residues: 1-1640 <HOL>
 A:Cross-references: UNIPROT:P04933
 C:Superfamily: major merozoite surface antigen
 C:Keywords: surface antigen

Qy	203	LNQFNQDELWVYLEKFFDLKALKANIRLOTADPSFEGKGNLVPFVYSFIRNPQONQKQWA	262
Db	168	IDGYEINELLYKLNIFYDLLRAKLN-----DVCANDYCOIPENLKIRA	211
Qy	263	SDLNQDQTVRLYLRTEFSPOAKTILDKYKQDETFLSSIDLKASNTSIPANENLDKQ	322
Db	212	NELDVLKVL-----FGYRKPL-----DNINKDN	234
Qy	323	LDVDLLDVSDFYGGQSETITSQVQKPVASERSLDRVKFKDQOKPRIEK-----PSLY	378
Db	235	VG-----KMEDYIKGNKTIENTNEL--IEBCKTIDKNKNATKEEEKKKLYQAOYDLSY	288
Qy	379	-----EYDALSFSYQOELVSKPNSIKOLVWAT-----	406
Db	289	NQLESEANHLISVLEKRIDTLKQENIKELLDKINEIKNPPANSNGTPTNLLDKNKIE	348
Qy	407	-----LARNLRFSLGKYNFLFDDLASHLDYVFLVSKAKIKQSSITKKLFIELPIKIS	458
Db	349	EHEKEIKETAKTKFNI-----DSLFTD--PLELEYL--REKNKN-----IDISAKVE	393
Qy	459	LKSSILGDOEPNKTILFEKEVTFKLNFRDVEIEKAFGLLYPGVNEELEQARK-----AQR	514
Db	394	TKESTEPNPN-----GVTPLS-----YNDINNALNELNSFGDLINP	432

Query Match 3.0%; Score 291; DB 2; Length 1640;
 Best Local Similarity 19.2%; Pred. No. 0.00015;
 Matches 322; Conservative 221; Mismatches 543; Indels 592; Gaps 79;

RESULT 18
A24594
probable major surface antigen (83K, 19K, 42K) precursor - malaria parasite (Plasmodium falciparum)
C:Species: Plasmodium falciparum
C:Date: 29-Aug-1987 #sequence_revision 29-Aug-1987 #text_change 09-Jul-2004
C:Accession: A24594
R:Holder, R.A.; Lockyer, M.J.; Odink, K.G.; Sandhu, J.S.; Riveros-Moreno, V.; Nicholls, S.; Nature 317, 270-273, 1985
A:Title: Primary structure of the precursor to the three major surface antigens of Plasmodium falciparum
A:Reference number: A24594; PMID:86014355; PMID:2995820
A:Accession: A24594
A:Molecule type: DNA
A:Residues: 1-1640 <HOL>
A:Cross-references: UNIPROT:P04933
C:Superfamily: major merozoite surface antigen
C:Keywords: surface antigen

Query Match 3.0%; Score 291.; DB 2; Length 1640;
Best Local Similarity 19.2%; Pred. No. 0.00015;
Matches 322; Conservative 221; Mismatches 543; Indels 592; Gaps 79;

Qy	203	LNOFNQDELWVYLEKFFDLEALKANIRLOTADFSEKGNLVDPVVYSFIRNPONQKEWA	262
Db	168	IDGVEEINELLYKLVFFDLLRAKLN-----DVCANDYQCQIPENLKIRA	211
Qy	263	SDLNQOQTVRLYLRTFSPSQATILKDYKYKDETFLSSIDLKASNGTSLPANENDLKQO	322
Db	212	NELDVLKGLV-----FGYRKPL-----DNIKON	234
Qy	323	LDVDDLVDVDFGGOSETITNSOVKVPVPAJERSLKDVRFKQOQKPRIEK-----FSLY	378
Db	235	VG-----KMEDYIKKNKKTIEINEL---IESKKTIDKNKATKEEEKKLQAOYDLSIY	288
Qy	379	-----EYDALSFSVQOELVSPNSIKDLVNAT-----	406
Db	289	NKOLEEAHNLSVLEKRIIDLTKKNENIKELLADKINEIKNPPANGNTPTNLLDKNKKIE	348
Qy	407	-----LARNRFSLGKYNFTFDLASHDYVFLVSKAKIKOSSITKKFLFELPIKIS	458
Db	349	EHEKEIKETIAKTIKFN1---DSLFTD-PLELEYVL---REKNKN-----IDISAKVE	393
Qy	459	LKSILGDQEPNIKTLFEKEVTFKLDNFRDVBIEKAFGLLYPGVNEEIQARK-----AQR	514
Db	394	TKGSTEENEPN-----GVYPLS-----YNDINNALNELNSGDLINP	432

Qy	515	ASPFKEKSKGGLKEPFSQOKBENSKAINNQBEGLEEDDNI	TERLPNSPIQVOQENAGLQAS	574	
Db	433	FDYTKPEPSKNITYD-----NERRKFFINE--	IKKKIKIEKKIESDKKSYE-----	475	
Qy	575	PDKPMIKDVONQRYILAKSIOIELIKAK-----	DYTKLAKLLGNRRHTYINISLRKLBOLF	629	
Db	476	DRSKSLNDITYKEKL-----LNEIYDSKFNNNIDLTN	FQWQMKGRVSYKV-----EKLIT	525	
Qy	630	DVNPRISSRDIKAKFVLDK-TENKYQWYIYSSASP	FQNKWSLFGYRYVLLGL--DPK	686	
Db	526	HPN-----TFASYENSKHNLKLTALKYKMEYD	SLRNIWVEKE-----LKYQKNLISKIENEI	578	
Qy	687	QTIHELVLGKAGLQFEGY-----ENLPSDENLE-	DLKNIRIKTLPFSQK-DNFKLSLL	739	
Db	579	ETLVENIKQDEQ--LFEKKITKDBENKPDKEIL	VSODIVRQVQVLLMMKIDELKKITQL	636	
Qy	740	DFNN-----YYDGEIKAPEFGLPLFPK-----	ELRNSNSNG	772	
Db	637	ILKNVELKHNIHVNSYKQENKQBPYYLIVLK	KBIDKLKVPMPKVESLINEKKNIQTQ	696	
Qy	773	GSQNSNPFWEQEIISQF-----		789	
Db	697	QSDNSPSTEGEITQCATTKPGQAGS	ALEGDSVOAQAEQKQAPVPVPVPEAKAQVP	756	
Qy	790	-----KQONLSNQD---OLAQF---STKIWEK-	IIGDENEFQNNRLQYLLKLDLQES	835	
Db	757	TPPAPVNNKTENVSKLDYLEKLYQLNTSYICH	KYLVSHSTMKEKILQYKIITYE-BES	815	
Qy	936	WINKTRNLWYTYIGDLKVKPKNNLEAKFROI---	SNLQELTAPYFYSAAALSNNWY	891	
Db	816	KLSSC-DPL-----DILLFIQNNIPVMYMF	DSLNNLSQLFWEIYEREMVCNLYKLK	867	
Qy	892	QDSGAKSTIIIEEIAELDPKVEKVGADVYOLK-		924	
Db	868	DNDKIKN--LUEEAKKYSTVSKTJSSSSMQPL	SJLSTPQDKPEVSANDDTSHSTNLNLSKL	925	
Qy	925	PHYAIGFDDNAGKFNQEVIRSSRTIYLTSGKS-		972	
Db	926	FENILSLGKNKNIYQELIGOKSENFEKILK	SDTFYNESFTNFVFKSKADDINSLANDES	985	
Qy	973	KNAPLGLGSFVLDYTERFGVFQKLATSLAVQH	KQK-EKTLPKKLNNDGTYLIHDKLKKPVI	1031	
Db	986	KRKKL-----EEDINKLKKTLQLSPDLYN	KYKLKLERLFDKKTATVGKYKM--QIKKLT	1037	
Qy	1032	PQISSPDKWFEGLKJNONGSQNV--NVSTFG	SIIESPYFESTMFQEDADLDQDQDSDR	1089	
Db	1038	-----LKEQLESKJNSLNNPKHVLFQNF	SFV-----FNKKKEAEI-----A	1072	
Qy	1090	QGNNSLNOBAGLLKQKLAILLGNQFTQYYQND-	KEIEFEIINVEK--VSELSFRVE	1144	
Db	1073	ETENTLENTKI-LLKHYKGL-----VKYNG	ESSPLKTSEESIQTEDNYVASLENFKVL	1125	
Qy	1145	FKLAKTILEDN---GKTIRVLSDETMSLI-		1176	
Db	1126	SKLSEGLKDNLNLEKKKLSYLSRGLGHLHIA	ELKEVINKNKYTGNSPSPNNVDVNNALESY	1185	
Qy	1177	---TPEMSAPPEVF-----DTKWVEQYDPR	TPLAAKTKFVLKFDQIPVDGSGNISDKWL	1228	
Db	1186	KKFLPEGEGDVAIVVSESGSDT--LEQSP	KKPASTHVG-----AESNTITTSQNVDD	1238	
Qy	1229	ASIPLVI-----	HQOMLRLSPVVKTIRELGL	1254	
Db	1239	DVIIVLIFGESEEDVDLIGVQVVTG	EAVTTSDINILSKIENEYEVLYLKLPLAGV	YRSL--	1296
Qy	1255	KTEQOQOQOQOQOQOQOKKAVRKEE	ELET--YNPKDEF-NITL-NPLTKAHR	LTLSNLVNN	1311
Db	1297	-----KKQLENVNMVTENVVKDILNSR	FNKRENFKNVLESDLIPYKDLTSSNVVK	1347	
Qy	1312	DPNYKIEDLKVIKNEAGHQHQLAFSLRANNI	KELMNTPIITFADYNPFY-----YNEDWR	1365	
Db	1348	DP-----YKFLNKEKRD--KFJSS	YNNYIKDSITDINFANDVLGVYKILSE	KYKSDLD	1398
Qy	1366	SIDKYLNNKGNVSHQOAAAGNOSGLIQRLN	KNIKPETPTPALIAALKDRNN--TWLSN	1423	

[illegible]

A:Molecule type: DNA	
A:Residues: 1-1979 <GAR>	
A:Cross-references: UNIPROT:O96133; GB:AE001375; GB:AE001362; NID:g3845105; PIDN:AA07181	
A:Experimental source: clone 3D7	
C:Genetics:	
A:Gene: PFB0145C	
Query Match	
Best Local Similarity 3.0%; Score 290.5; DB 2; Length 1979;	
Matches 293; Conservative 301; Mismatches 559; Indels 563; Gaps 73;	
QY	248 VYSFIENPQNEKAWSLDQDQKTIVLYLTRETSPOAKTILDKYKDETFLSSIDLKAS 307
DB	2 VFTF-KNKKKKKKAASD-----KVSKESENEEDNENNEKESDSWYKLIETKCK 51
QY	308 NGTSLFANENDLKQDLVDLLVDSYFGGQSETITSQVQKVPASERSLKRVPFKDQ 367
DB	52 SKTK-YKDNLSLDNINEDIINNNNNNNNNNNNDN-----NNDNNNNNNNNE 104
QY	368 QKPRIEKFSLYEYDALSFYSQQLVSKPNSIKDLVATL--ARNLRFSLGKYNFLPDDL 425
DB	105 NNDNNNNFNYS-DEIS-----KNIIHKONELNQLKDTLKSISLNKIVNYESKIBEL 158
QY	426 ASHL-----DYFLVSAKIKQSSITKFLFELPIKISLSSILGDQEPNI--- 471
DB	159 EKELKEVKDKINDDY---ENKLKEKEDFVKQIDM-----LNKENLLQELDLNKR 210
QY	472 -KTLFEKE--VTFKLDNFRDVEIEKAFGLLYPGVNEBELEQARAKAASFEKESKKGLK- 527
DB	211 EKKINEKKNIIKKEETFNIE-----KEYLEKNKERETISIEIIDIKKHLEK 258
QY	528 ---EFSQKKE---NSKANNOGULEE-----DDNITERLPENSIQVQ 565
DB	259 LKIEIKKEDLENLKNLLSKENLVKELKGCVKERNETINSLNNDNIIIEKKYKULVEY 318
QY	566 QENAGLGAS-PDKPYMKQVQNYRYLAKSOIELIKAKDYTKLAKLSNRHTYINSLRL 624
DB	319 LEEKNQIDLLNKOKEKEKEKER---EKEKEKEKEKEKYDTYILIKELDS-----KISI 370
QY	625 KEQLFDVNPRIPSRDIETAKFVLDTKTKNYQYIYSSAPVQNKWSLFGYYRYLLGLD 694
DB	371 LEKHSIKVR---EMDIEKEE-----388
QY	685 PKQTHLVLKQKAGLQFPGYENLPDSFNLDELKNIRIKTPIFQKDNFKLSLLDPNNY 744
DB	389 -----HNFHMEQDLKDKNSFV-----KNNQLKV-----414
QY	745 YDGEIKAPEGLPLPLPKELRRNSSGGSGNSPWEQBIISQFDONLSNODQLAQS 804
DB	415 YKCEIKN-----LXTELEKEKEKEDJENSVKEBINKLIMQLEK-----EKQILAFN 462
QY	805 TKIWEKIIGDENEPDQNNRLQYKLLKDLQBSWINKTRDNLWYTYLGDYKLRPKNNLEAK 864
DB	463 KNHKEEIHGLKEELKESVKIKTIETQELQEM-----VDIK 497
QY	865 PROISNLOELLTAPYTSAL-----SNWNVYQDSGAKSTIIFEEIAELDPKVEKVGAD 919
DB	498 QKELDQLEQKYNQAQISISIELSKKEEYNOY-----KNTYI-EEINNLNKL-EETNKE 550
QY	920 VYQLKPHVATGPDNACKPQOEVRSSRTIYLTCKSKLEADTIDOLNQAQVNAFLGL 979
DB	551 YTNLQNNY-----TNEINMLND-IHMLNGNITKMTQI 593
QY	980 QSFYLDTRFVGFKLATSLAVQKQKEKTPKPKLNNDGYTLIHDKLKXPVIFQISSPE 1039
DB	584 STLKNDV-----HLLNEQI--DKLANNEKGTU-NSKISELNQVIMDLKEE 624
QY	1040 KDWPEGLK-----NQNGQSQVNVSTFGSIIISPFSTNFOE 1076
DB	625 KDFLNNQIVDLNSQIDLLTRWEEKENKMLQENKYQSBELLR-GNIKSENILNNDDEE 693
QY	1077 DADLDQDQDSDRSQNNSLDNQEAGLLKQ---KLAILGNQFOYQYQNDKEIEFBIIN 1132
A:Superfamily: surface-located membrane protein lmp3; tetratricopeptide repeat homology	

RESULT 22

JC6009

surface-located membrane protein lmp3 precursor - Mycoplasma hominis

C:Species: Mycoplasma hominis

C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 09-Jul-2004

C:Accession: JC6009

R:Jadefoged, S.A.; Jensen, L.T.; Brock, B.; Birckelund, S.; Christiansen, G.

J. Bacteriol. 176, 2775-2784, 1996

A:Title: Analysis of 0.5-kilobase-pair repeats in the Mycoplasma hominis lmp gene system

A:Reference number: JC6009; MUID:96213016; PMID:8631664

A:Accession: JC6009

A:Molecule type: DNA

A:Residues: 1-1302 <LAD>

A:Cross-references: UNIPROT:Q49547; EMBL:X95601; NID:g1197335; PIDN:CAA64858.1; PID:g1197

C:Genetics:

A:Gene: lmp3

C:Superfamily: surface-located membrane protein lmp3; tetratricopeptide repeat homology

[illegible][illegible]

A:Residues: 1-2829 <GAL>			
A:Experimental source: Belem strain, merozoites			
A:Note: sequence extracted from NCBI backbone (NCBIN:108114, NCBI:P:108115)			
Query Match	2.9%; Score 282.5; DB 2; Length 2829;		
Best Local Similarity	19.4%; Pred. No. 0.00075;		
Matches 423; Conservative 328; Mismatches 757; Indels 673; Gaps 111;			
QY	49 LIDSVAPK-----STANFTSDYQSVKALLNGKTPD-----PKSSEPTDFVSKPD 94	Db	1524 SIVAVTQDNLNFESLMKEVKVCKEKK--NDEAEKYSALKPYD--GRIKARVSEN--E 1577
Db	641 ILDKMAKVHYLLKELLSLGKGSVYFTENNELNTASYDNMEGFAKKEKADNDINALYN 700	QY	835 SWINKTRNLVTVILGDKLKVKK-----NMLEAK-FROISNLOELLTAPTYS--AALSNN 887
QY	95 -----FLTNNGRTVLEIPKKYQVVISSEPSDDKERR-RLGFLHLEKLE 137	Db	1578 RKISE-----LKEKAKYKESKSSQNDVSTKSLQIDNCRCQDLSVLNIGRVKON 1628
Db	701 SVYREDINALIEVEKFTVENKESTLEMLKDEM---EEKQDAKETFAKLVFSDDKLT 757	QY	888 WNYTQDSGAKSTIIFEEIAELDPKV---KEHVGADVQVLKPHYAIGFDNAGKFNQEVIR 944
QY	138 DGNIAQSATKFTYLLPLDMPKALGOYSIVDNKFNFLIIH-PLSNFS-AQSIKPLALTR 195	Db	1629 ALQVFDSDAKSMKSVLPISLGAESKSLDKVKAESYEK-----NLETVQ 1673
Db	758 DVYTKMSAE-----VTNAGIKKEIAQKQFEN--VHKMKFEFSAFTKFEALQN 805	QY	945 SSSRTIYLTSGGKLEADTIDQLNQAIVK-----NAPLGLOSFYLDTERF 989
QY	196 S-----SDFTAKLQFNQDELWYLYLEKFPDLEALKANI-RLQTAQ-----FSF 238	Db	1674 NEMSRINVEEGLTDIDKKITDIEIDLKMKQVBEGLLQIKENADRKSNFELVSGEI 1733
Db	806 SMQOVNQEGBDAIEKKHQRSEK-----BEYFKNESVEEDLSRETEQEYTKHKNFNR 860	QY	990 GVFOKLATSLAVQHKQKEKTLPKKLANNQGYTL--IHDKLKKPVIPIQISSPEKDFEGKL 1047
QY	239 EKGVLDPFV-----YSFIERNPONQEWASDLNQDQKTVRLYLRL 277	Db	1734 NALLDPSTSIPIKULKEYDMTGDLYNGVKNNEIHGEFTK-----1774
Db	861 RKGEISAEITNMREVINKIESQLNYGVYIEKYFSLIGDQNEVSTAKALK--EKIVSDSLR 918	QY	1048 NQNGSQNVNVSTFGSIIESPVFTNFOESDADLDQDQDSDRQGNNSLDNQEAGLLKQKL 1107
QY	278 TEPSPQAKTILDKYKDETFIASSI-----DLKASNGTSLFANENDLKQDLVDLDD 329	Db	1775 -----SYNLIETHLSNATD--YSVTFEKAQSLRE-----1801
Db	919 DKID-QYETFEKTEKSAVENTVSTIQLSKAIDSLKRLNGSI---NNCKKYNTDIDL- 972	QY	1108 AILLGNQFIQYQQNDKEIFEIINVEKVSLSFRVFP--KLA-----1148
QY	330 VSDYFGQSBETITNSQV-KVPVASE-----RSLKDRV-KPKDQKQKPRIKFP 375	Db	1802 ---LAEKEEHLRRREBEAIFLLNDIKKVSLLKEMMKVKVSAEYEGMKRDHTSVSQLV 1858
Db	973 -----RSKIKTLREEVQEMKRGDKCGENTTALLKSLRDKMGKINEKLNKDLGRLNSL 1025	QY	1149 ---KTLENGKTIRVLSDETSLIVNTTIEKTPEMSAVPEVDTKWE-----QYDPR 1198
QY	376 SLYEYDALSPYQELVSKPNSIKDLVNATLARNLRFPSLGKYNFLPDLASHLD---Y 431	Db	1859 QDMKTIIVDELKTLNDISE--CSSVLNNV-----SIVKVKESKHADYRRDANSMTESM 1910
Db	1026 DTKKEDLLKFPYS---ESKSIHLKSD-----QKGPQDPLNRID-EWEIDIKRDVDELNVN 1075	QY	1199 TPLA-----AKTFVLKPKDOIIPVDGSGNISDKWLASIPLVIIHQMLRLSPVVKTI 1251
QY	432 YFLVSKAKI---KOSSTTKKLFTELPIKISLKSILGDOEPNIKTILFEKEVTFKLNDFRD 488	Db	1911 VTLANYFLSDEAKISGMEFNAEMKSNFKTDLE-----LEIFSVISNSNE 1955
Db	1076 YQVISENKVTLFNKNSVT---YIEA-----MHSHINTVAHGITSNKEILKS 1119	QY	1252 LGLKTEQEQQQQ--OQQOQQOQQPOKKA-----VRKE--EELETYNPKDEF-----NI 1293
QY	489 V-BIEKAFGLL-----YPGV-NBELEQARQAQASFEKEKS--KGLKGFSSQK-----533	Db	1956 LLKKIEQDSNDVIQKERESEQLAKDQATDIYVNIKLNKEFNEKLEAKNEKEVSEKVEA 2015
Db	1120 VKEVEDKLVLEQVEDYKYNPENENKQLEAIRGSMKLEKVINKHVSEMTQLESTANTL 1179	QY	1294 LNPITKA-----HRLTSLNVLNNDPNYKIEDLKI-----KNEAGDHQLAF 1334
QY	534 BENSKAINQOEGLLEDD-----NITERLPENSPIOYQOEN-----568	Db	2016 LKRLSQVEGIRCHFNHFRL-----DNTBELNLKQMTVIYRDKKSERESGLQEMEN 2068
Db	1180 KSNAGKNEHDLLELNKTKGQMRDIYEKJKAELKEGTVELKDANEKANKVPEPE 1239	QY	1335 SLR--ANNIKRLMNTPIPTFA---DYNPPFYNYEDWRSIDKYLNNKGNVSSHQOQAAGN 1388
QY	569 -----AGLGASPKPYMIKDQVQRYVLAKSQIOELIKAKD-----YTKLAKLL 612	Db	2069 EMWTTYSNSTQLEGIIVSAGESKEDIKLEKRENEEMNISE-----KLTSDSKVIEMN 2122
Db	1240 RNIIGHVLERITVEKDAGKVVEEMS---LTKIEKLIQETSDDSQNELVTTISITKHL 1295	QY	1389 QGSLGIQRLNKNIKPETFTPALIALKDRNNTNLNYSOKIIMI-KPKYLVERSIGV--PW 1445
QY	613 SNRHTY-NISLRLEQQLFDVNPRIIPSSRDIEKAFVLDTKTKNRYQWYSSASPVFNQKW 671	Db	2123 STIDELYLKGNQCAHWI--SLISVTANMKT-----SKCLIMINKENTKCVYIKDN 2175
Db	1296 ENAKGYEDVIKRNEEDSIQREKAKSLTLDENKKLVQVVMNLQSAIQGNAG-ISKELN 1354	QY	1446 STGLDGYTQSEOTKDTSSSSQOQGFQDQFIQALGLKNTXY--HGKLGLS-----IR 1495
QY	672 SLFGYRYLGLD-----PKQTIHELVLKQKAGLOP---EG-----Y 706	Db	2176 SSSTDGYV---ETLKGFGYSGKLTFFSSASEIVQADTVSYNFAKHEKESLNADRIKKELY 2232
Db	1355 ELKGVIELLSTNYSSTILEYVYKGNSSSVRFPSQLANGFEFTKAEGEKNASARLAEBKLK 1414	QY	1496 IFDPGNELAKIKDASNKGEELKLSYD-----LFKNYLNEYKK-----SPKIA 1540
QY	707 ENLPSDNLDEL---KNIR-IKTPFQSKDN-----PK-LSLLDFNNYDGEIK 750	Db	2233 LFHONSDDISIVEG-----GVQNMALYDKLNEKREMDLYRN-ISETKLKQHEHSTDVDF 2286
Db	1415 EQIVKDLJDYSDIDDKVKVIGIKREILUKMESALTFWEESEKFKQMSCHWENAKGKKK 1474	QY	1541 KGMTNIHPDQKQVYPPNPKLPENYLNVLNQPWKVTLNYSDDFITNLFVEPEGSDRGSGT 1600
QY	751 APEFGFLFLPKELRRNSNGSGQNSNSPWQ-----E 784	Db	2287 KPMIELHKGWNETNNKSLLEKEKLSVNDHMSM---EAEMIKN-----GL 2330
Db	1475 I-----EYLKXNGDGKANITDSQWEEVGNVSKAEHAFHTVEAQVQKTAFCPE 1523	QY	1601 KUKQVLOKQVNNYA-----DWGSAYLTFWYDK---NLIITQPNVITANIA 1643
QY	785 IISQF--KQNLNSQDQLAQFSTKIWEKIIGDENE-----PDQNNRLQYKLLKDLQBE 834	Db	2331 KYTPESVQNNINIYSYIEAEVKTLEBIDRDYDGNQYQVVEEHKKQFSILIDRTNLMDDI- 2389
		QY	1644 DVFIKDVKELEDNTKLIAPNITQWPNISGSKEKFKYKPTVFFGNWENENSSMNSQAQTPT 1703
		Db	2390 EIPFK-----ENNYNLMEVN--TETIHRVNDYIEKIITKLV-----QAKT-E 2428
		QY	1704 WEKIREGFALQALKSSFDQKTRFTFVLTTNAPLPLWKYGLPGFQNGFNPFTQDWRLLVFOND 1763

Db 1296 LKQRIVDKRLKLNERNLHLSNTERKRAVLDEQISYFVKORQATDAIIASHKE 1355
QY 1525 FKNYLNEYKGPKEAKGWTNIHPDOKEYPNQKLPENYLVNQPWKVTLVNSDDPI 1584
Db 1356 VKKGEGLQKLVLETRKTLNNDPAKSRQREBFENQRLKLELQKLTQOTQSSNPFK 1415
QY 1585 TNLVFEPEGSDRGSGTKLVQKQVQKQVNNYADWGSAYLTFWYDKNIITQPNVITANAD 1644
Db 1416 TRAIQIENSYKRGMBELN--FQKE--FDKN-----KSLRYE 1449
QY 1645 VPIK---DVKELEDNTKLIAPNITQWPNI---SGSKERFYKPTVFFGNWENENSSMNSQ 1698
Db 1450 YFRKMRDEIRKESQVKVLKE--TQKANLLEAQAQKLNIEKNTIDFKKE-----1499
QY 1699 AQPTWEKIREGFALQALKSSFDQ 1722
Db 1500 -----LKAFKDKVDQ 1509
RESULT 27
D71623
erythrocyte membrane protein pFEMP3 PFB0095c - malaria parasite (Plasmodium falciparum)
C:Species: Plasmodium falciparum
C:Date: 13-Nov-1998 #sequence_revision 13-Nov-1998 #text_change 09-Jul-2004
C:Accession: D71623
R:Gardner, M.J.; Tettelin, H.; Carucci, D.J.; Cummings, L.M.; Aravind, L.; Koonin, E.V.;
Pextea, M.; Salzberg, S.; Zhou, L.; Sutton, G.G.; Clayton, R.; White, O.; Smith, H.O.
Science 282, 1126-1132, 1998
A:Title: Chromosome 2 sequence of the human malaria parasite Plasmodium falciparum.
A:Reference number: A71600; MUID:99021743; PMID:9804551
A:Accession: D71623
A>Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-2441 <GAR>
A:Cross-references: UNIPROT:O96124; GB:AE001371; GB:AE001362; NID:g3845092; PIDN:AAC7180
A:Experimental source: clone 3D7
C:Genetics:
A:Gene: PFB0095c
Query Match 2.9%; Score 281; DB 2; Length 2441;
Best Local Similarity 20.6%; Pred. No. 0.00071;
Matches 349; Conservative 234; Mismatches 614; Indels 498; Gaps 85;
QY 74 LNK-----TFDP-----KSSEFTDFVSKFDPLTNNGRTVLIEIPKQYQVISEFS 118
Db 481 VNGKCLKSTNPFKRRNKLKERKQELHKFNKY-----KKYQKLE---523
QY 119 PEDDKERFRLGFLHKEKLE--DGNIAQSATKFYLLPLDMPKAAALGOYSIVVDKFNFLI 176
Db 524 -----REKRENPGE-----PLNTPFIHVIRPSDLMDKGENKSA 557
QY 177 IHPLSNPSAQSIK-----PLALTR--SSDFIAKLQNFNQ--DELWVYL 216
Db 558 GHFPKPYQPTKGLYEBSHVAKDYOLEHEPPTKLPEYKGVHSREYQLDNEVRDELPEYE 617
QY 217 EKFPDLLE-ALKANIRLQATDFSEKGNLVDPFVYSPIRNPQOKEWASDLNQDKTVRLY 275
Db 618 KGHVSREYQLDNEVRDELPE--YEKGVH-----SREYQLD-NEGFPSTLKEY 660
QY 276 LRTEFSPQATILKDYKYKDETFELSSIDLKASNGTSLPANENDL--KDQLDVLDDVSDY 333
Db 661 DQTELAKGKDITNKPHEVSDEYDQSEL-----AKGKDI TNKPHEVSDEYDQTEL 709
QY 334 FGQSETITNSQVQKVPASERSIKRVPKQKQKPRIKFSLEYEDALSFYSQLELV 393
Db 710 AKGEVTKNKHENLE-----EYNETDLAKGEVTKNKH-----ESVDEYDQ--SELAKGKDI 760
QY 394 SKPN-SIKOLVNATLARNLF-----SLGYNFLPDDL-----SHLDYFLVS 436
Db 761 NKCHESVDEYDQTELAKGEVTKNKHENLEYN--ETDLAKGEVTKNKHENLEYN 818
QY 437 KAKIKQSSITKULFIEL-----PIKISLKSSILGDQEPNKTLP-----KEVTFK---482

Db 819 LAKGE--VTNKAHENLEYNETDLAKGEVTKNKAHENLEYNETDLAKGEVTKNKAHEN 876
QY 483 LDNFRDVEIEKAFGLLYPGVNEELEQAKAQRASPEK--EKSXKGLKEFSQO-----KEE 535
Db 877 LEEYNETDLAKGEVTKN--ARENLEYNETDLAKGEVTKNKAHENLEYNETDLAKGEV 935
QY 536 NSKAINNOEGLEEDD-----NITERLPENSPIYOQOENAGLAGSPDKPMIKDVQNO-RY 589
Db 936 TNKAHENLEYNETDLAKGEVTKNKAHENL--EYNETDLAKG-----KEVTKNKA 985
QY 590 YLAKSQIOELIKAKDYTKLAKLLSNRHTYN-----ISLRLKEQLFDVNPIPSRD 640
Db 986 NLEYNETDLAKGEVTKN--ENLEYNETDLAKGEVTKNKAHENLEYN-----ETD 1038
QY 641 IEKAFVLDKTEKN--KYWQIYSSASPVFQNKWSLFGYRYLLGLDPPQTHIELVKLGOK 698
Db 1039 LAKGEVTKNKAHENLEEYEEKDYMKNNELQNGS-----DGLKENAELKNK 1084
QY 699 AGLOPEGYENLPSPDNLEDLKNIRIKTFLPSKONFKLSLLDFNN-----YVDGEIKAP 752
Db 1085 B-LRNKSGDLKENAELKN--KELRNKGS--DGLKENAELKNKELQNGSGLKENAELKNK 1141
QY 753 EF-----GL-----PLFLPKELRNSSNGSGSQNSPWEQEIISQFQDNLSNQDLAQF 803
Db 1142 ELQNGSGLKENAELKNKEL--RNKGSGLKEN-----AELKNKELQNGS-----1186
QY 804 STKIWEKIIGDNEFDQNNRLOYKLLKDLQESWINKTRDNLTYLGDYKLVKPKNNLEA 863
Db 1187 -----EGL--KENAELKNKELQNGSGLKENAELKNKE-----LQNGSGLKENAEL 1233
QY 864 KFRQISN--LOELLTAFYTSAAALSNNMNYOOSGAK-----STIIPEETAELDPKVKRQV 917
Db 1234 KNKELRNKSGELKENVYNNNDLKNNDIQNDLSNKMKNKELLANKDISNKMKNKELLN 1293
QY 918 ADVYQLKFHYATGFDNAGKFNQEVIRSSRTIYLTSGKSKLEADTIDQLNQAQVKNAPL 977
Db 1294 KDL-----SNEDMKNELLNKDIRNKDLKSIG-----NMEQNTGLKNTP- 1333
QY 978 GLQSFYLDTERFQFKLATSILAVQHKEKTLPKKLNNDGYTLIHDKLLK--PVIPQISS 1036
Db 1334 -----SKGOQNTGLKNTPNERQNTG-----LKNTPSEGOQNT 1366
QY 1037 SPEKDFEGLN-----QNGOSQVNVSTFSGIIESPYFSTNFBQSDALDQ--DGQDD 1087
Db 1367 GLKNTFSEGOQNTGLKNTPNERQNTGLKNTPS-----EGQNTGLKNTPIEGOQN 1417
QY 1088 SRQGNNSLDNQBAGLLKQKLAILLGNQFTIYYQONDKEIEFBIINVEKVSLSFRVEFKL 1147
Db 1418 TGLKNTFSEGOQNTGLKN-----TPNERQNTGLKNAANKGOQNTGLKNTP 1463
QY 1148 AKTLBNGKTIIRVLSDETHSLIWNITIEKTPMSAVPEVDTKWVEYQDPTPLAAKTF 1207
Db 1464 SKGOQNTG-----LKNTPNERQNTGLKNTP-----NERQNTGL-----1498
QY 1208 VLKFKDQIPVDGSGNISDKWLASIPLVIHQMLRLSPVVKTIRELGLKTEQOQOQOQO---1264
Db 1499 -----KNTPSEGOQN--ND--LKNTPNERQNTGLKNKNTASKGOQNTGLKNAPNERQNTGL 1550
QY 1265 -----QOQOQOQOKKAVRKEELET--YNPKE-----FNILNPLTKAHLITLSNLVNDPN 1314
Db 1551 KNTFSEGOQNTGLKNTPSEGOQNTGLKNTPNERQNTGLKNKNTASKGOQNTG--GLKNAPN 1607
QY 1315 YKIEDLVKVNAGDHOLAFSLRANNIKELMNTPTIPADYNPPFYNNEDWRSIDKLYNNK 1374
Db 1608 ERQQNTG--LKNTPSEGOQNTGLKNKNTASKGOQNTGL-----KNTPSEGOQNTGL 1654
QY 1375 GNVSSHQQAAG-----GNOGSLIORLNKNIKPEITFTPALIALKRNNTNLSYSDK 1427
Db 1655 KNAPNERQNTGLKNTPSEGOQNTGL-----KN-----TPS-----EGQNTGLKNTPS- 1698
QY 1428 IIMIKPKYLVERSIGVPSWTGLDGYIGSBQTKDQ-----TSSSSQOKFQDQFIQALGKNT 1484

Db 1699 -----GGQONTGLKNTPNRQONTGLKNTPSEGOQN-----TGLKNT 1735
Qy 1485 --EVHGKGLSIRIPDPGNEALAKIDASNGKGEKLLKSYDLFKNYLYNEB-----KKSP 1537
Db 1736 PNERQONTGL-----KNAANGQONTGL-----KNTPNRQONTGLKNT 1775
Qy 1538 KIAKGTWNIHPDQKPEYPNQKLPENYLNVLNQPWKVLYNSDDFITNLFVPEGSDRG 1597
Db 1776 SEGOQNTGL-----KNTPSEGOQ-----NTGLKNT-----PSEGOQNTGLKNTPNRQ 1818
Qy 1598 SGTKLKQVIOQVNV 1612
Db 1819 QNTGLKNAANGQON 1833

RESULT 28
S73852
hypotheetical protein MG218 homolog F10_orf1818 - Mycoplasma pneumoniae (strain ATCC 29342)
C:Species: Mycoplasma pneumoniae
A:Variety: ATCC 29342
C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 09-Jul-2004
C:Accession: S73852
R:Himmelreich, R.; Hilbert, H.; Plagens, H.; Pirkl, E.; Li, B.C.; Herrmann, R.
Nucleic Acids Res. 24, 4420-4449, 1996
A>Title: Complete sequence analysis of the genome of the bacterium Mycoplasma pneumoniae
A:Reference number: S73327; MUID:97105885; PMID:8948633
A:Accession: S73852
A>Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-1818 <H1>
A:Cross-references: UNIPROT:P75471; EMBL:AE000051; GB:U00089; NID:g1674211; PIDN:AAB9617
A>Note: the nucleotide sequence was submitted to the EMBL Data Library, November 1996
C:Genetics:
A:Genetic code: SGC3
C:Superfamily: Mycoplasma genitalium hypothetical protein MG218

Query Match 2.9%; Score 280.5; DB 1; Length 1818;
Best Local Similarity 19.6%; Pred. NO. 0.00049;
Matches 310; Conservative .261; Mismatches 628; Indels 383; Gaps 66;

Qy 56 KPSTANFTSD-----YQSVKALLNG-KTFDPKSSSEFTDFV--SKFDFLTNNGR 102
Db 323 KASLANLTKEKERLSAEKDSFERURNTALDNINRMEQENALFAKHLEQQQYEFERKQES 382
Qy 103 VLEIP---KTYQVWISPEPDDKERFRLGFLHKELE-----DNTAQSATKFIYLLPL 154
Db 383 LKLETEHQLOKRGIFEKISEAKSEALLQERELLEKRREIDDLTQASLVEYEQRR 442
Qy 155 DMPKAALQGYSYIVDKNNLIIHPLSNFSAQSIKPLALTRSSDFIAKLQFNQDELWV 214
Db 443 NQ---VLKEKHQVQQHPQL-VHAKK-----LDQKRHYLAQKRI---DEEQI 485
Qy 215 Y-----LEKFF-----DLEAKNIRLQTADFS---PEKGNLVDPF 247
Db 486 FLKEKIAATERREKLYLVKKQDKQKENDLLIFEKOLROYQADFENEIEKQ--ELF 543
Qy 248 V-----YSFIRNPQ---NQK-----EWA-----SDLN-----QDQKT 271
Db 544 ASQKSLQKSFQTKNKEAELNKAQKIAEDNAHLKQNKHHADLEIFLEGFNHLQKEKH 603
Qy 272 VRLYRTFSPOAKTILDKYKDBTFL-----SSIDLKASNGTSLFANENDLKQDLVDL 327
Db 604 KLEARTQFDNRVLSLARSFKQKQAEKVQKQSLTAAPFNKQSEBAVERDWMKRL--ANL 662
Qy 328 LDVSDYFGQSETITSNQVVPASERSLDRVKFKDQKQKPRIKPSLVEYDALSPYS 387
Db 663 EKQKEMLDGVHQVDFENSLNISKLAEREL--ATKFEKELEAAQKQSLDNNNNAGL 720
Qy 388 QLOELVSPNSIKDLVATLARNLRFSLGKYNFLPDDLIASHLDYVFLVSKAKIKOSSITK 447
Db 721 QLDKLSLKTLELEASKERILDF-----YDESSRIADYESDLQARLAEVKTLE 772
Qy 448 KLFIPLPIKISLKSSILGDQEPENITLFEKVTFKLDNFRD-----VEIEKAFGLLPGV 502

Db 773 K-----NQETAAKS-----BRELKVALEKLNQAKKAFQIRKQQOLEIASV 814
Qy 503 NEELEQAR---KAORASFEK--EKSXKGLKEFSQOKEENSKAINN---OBGLEDDNIT 553
Db 815 KQQAQKANLLKQQAELDKQTELEAAFLQDDTKKELEKALHSVSKQBELLEERSFL 874
Qy 554 ERLPENSPIQYQENAGLGASPKPYMIKDQVQRY--YLAKSQIOELIK-----601
Db 875 -----LQKQREFAEHVAGFKQVHFKTQWRLSEFNKQOQSEQIKRETELKIAFAD 926
Qy 602 -ADYTKLAKLLSNRHTYINSLRKE-----QLFDVNPRIPSSRDEIAKAEVL 648
Db 927 LKKDY-QLFELQKNQEFQOIEQKHLELLAQKAEKQELQEKATALASQDQDTVAQL 985
Qy 649 DKTEKNKYWOIYSSASPVFONKWSL-FGYRYLLGLDPKQTIHELKVLGOKAGL-----701
Db 986 DLARQOHELELRQNAF-----NQASLSLNKQREQLTNQVKVLHGELKKRHEKLTLDKRLLA 1041
Qy 702 -----QFEGYENLPDFN-----LEDLKNIRIKTLPFSQKDNFKLSLLD 740
Db 1042 EKEKDHKKDAEINOREFKQFENEYADFOAKRELBELQNLQIR-----RNLEQSNASILK 1095
Qy 741 FNNYDGEIKAPFGLPLFLPKELRNSSNGSGSONSPWEQEIISQFKD-QNLSNQ--797
Db 1096 KRN-----QLTLDPALLRKVQHNTQTRVOLNTQI---KEFLLEKGNFQKASDEAA 1143
Qy 798 -----DQLAQFSTKIW---EKIIGDENEFQDNRLQYKLLQDLQESWNTKTRDNLWT 847
Db 1144 LQKALLIKRLRSFASKLOQREALAIQKLFDRD-----EQKSEINNAKLQ---1192
Qy 848 YLGDKLKVKPKNNLEAKFRQISNLQELLTAFYTSAAALSNNNWYQDSGAKSTIIFEEIAE 907
Db 1193 ---EQFKLEKQNFDEAKOKLIEBFQDQRLDVEKRLKQKLQVQLKSLSYLYTKNRAD 1249
Qy 908 LDPKVKKEGVADYVQLKEHYAIGFDDNAGKFNQEVIRSSRTIYLTSGKSKLEADTIDQ 967
Db 1250 LSQOQLQHKYANLLELK-----EKLTAKRALDKGHRAIYGVKWAQFVSE 1293
Qy 968 LNAVKNAIPLGLOSFYLDTER-FGVFQKLAISLAVQHKQKEKTLPLKLNNDGYTLIHDKL 1026
Db 1294 LRQEKQJL-LSAQKQVDDKSLLEQNGHQHQLNLSSETKKRQSLHEDINK-----FDQR 1346
Qy 1027 KXPVIOISSPEKQWFEGLNQ-NGOSQNV--NVSTFGSIIESPYSTNPQEBADLDQD 1083
Db 1347 RKEAVSILNSHK-----KLKQKEGELQGLQKLSLKTQIEQE-FSKLYQOREKLDQ 1399
Qy 1084 QODDSROGNSLNOEAGLAKQKLAILLGNQFIQYQONDKEIEFEIINVEKVSLSFRV 1143
Db 1400 RTLLSKLHRELKAQNEATAHKNREVLEIN----YKKELQRLTEKSEFDNNKRNLFY 1455
Qy 1144 EFKLAKTLENGKTIRVLSDETMSLIVNTTIEKTPEMSAVPEVDTKWVEQYDPTPLAA 1203
Db 1456 FRKIRNEIEKKAHIKTVLEETQK-----KRLHVEAVLHLQKQSIIS 1500
Qy 1204 KTFEVLKFKQOI PVDGSGNISDKWLASIPLVHQQMLRSLPVVKTIRELGLKTSQOQQOQ 1263
Db 1501 KGQELKEIKERSVSDISH--TNKQREBELSLLHQKLLQKNLAEREINNNKDSLLTQKI 1558
Qy 1264 QQQOQQQPOKKA-----VRKEEB-----LETYN---PKDEFNINLPIITKAHL 1303
Db 1559 QTAQKQLSEKARILKLEKRAVEQQYQAEITLKTNRADLEKNDKHLFPFLFKIN--1616
Qy 1304 TSLNVLNVDPNY-----KIEDLVTKNEAGDHLAFSLRANNIKELMNTPTIFFA 1352
Db 1617 -----GNDMNYPPYPWPFPQKQED-----SSNQIRHLFEQQLQFM 1653
Qy 1353 DYNPFFTYNE--DWRSTDVKYLNK-KNVSSHQQAAGNQGSGLIQRLNKI-----KPE 1404
Db 1654 QQR---YENELTELRRORALLEKLDQIQLSLSAKKNDPEKVEQMOKLLEKTEQKLS 1710
Qy 1405 TFTPALIALKDRNNTNLNSYSD 1426

Db 1896 KQDLFKK-LNRGGVRVLIGSPAKMGVG-TNVQERLVAMHEDCPWRPDELLQMEGRGIRQ 1953
Qy 1553 YPNPNOKPLNYLVNLQPNKVTLYNSDDITITLNFVEPEGSDRGSGTKLQKQVQKQVNN 1612
Db 1954 GNIHLQNDPENF-----RMKIYRYA-----TEKYDSRMWQIETK-SK 1991
Qy 1613 NYADWGSAY---LTFWYDKNIITNPVNITANIA--DVFIKDVK---ELEDNTKLIAPNI 1664
Db 1992 GIEQFRNAHLGLNELEDFNMGSSNASMKAEGATGNPLIIEVVKLRAEIKSEE----- 2044
Qy 1665 TQWPNISGSKEKYKPTVPFGNWNENENSMNSQAQPTWEKIREGFALQALKSSFDQKT 1724
Db 2045 -----SKYKAPNKEHYF-----NEESLKNASKL-----DYLKQELK-DLETLQ 2082
Qy 1725 RTFVLTNNAPLWKYGLPGFQNGPNFKTQDWRLVFOVNDNQIAALKRVQEODRPEKSSD 1784
Db 2083 RSVIIPHTHEIKLDL-----KNEESKDYELIKVEVEPLENASMSELTHKKLKE 2134
Qy 1785 KDKQWIKFKVVIPEEMPNSGNIRFVGMQIQGNTLWLPVINSVIYDF--YRG----- 1837
Db 2135 QNKQ-----IAEQ--NKEKDAIKQFASNLNTLF---VNEEDYKLEYGFVVNA 2181
Qy 1838 -----TGDSNDVANL--NVAPWQVKTIAFT 1860
Db 2182 YKTRYQVEFSLSPKIDIPNIAYSLAIFKTLTST 2215

RESULT 30
C90538
Hypothetical protein MYPV 2110 [imported] - Mycoplasma pulmonis (strain UAB CTIP)
C:Species: Mycoplasma pulmonis
C:Date: 24-May-2001 #sequence_revision 24-May-2001 #text_change 09-Jul-2004
C:Accession: C90538
R:Chambaud, I.; Heilig, R.; Ferris, S.; Barbe, V.; Samson, D.; Galisson, F.; Moszer, I.;
Nucleic Acids Res. 29, 2145-2153, 2001
A:Title: The complete genome sequence of the murine respiratory pathogen Mycoplasma pulm
A:Reference number: A99512; MUID:21267165; PMID:11353084
A:Accession: C90538
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-3216 <KUR>
A:Cross-references: UNIPROT:Q98029; GB:AL445566; PID:g14089624; PIDN:CAC13384.1; GSPDB:G
A:Experimental source: strain UAB CTIP
C:Genetics:
A:Gene: MYPV 2110
A:Genetic code: SGC3

Query Match 2.9%; Score 279.5; DB 2; Length 3216;
Best Local Similarity 17.6%; Pred. No. 0.0012;
Matches 412; Conservative 361; Mismatches 824; Indels 743; Gaps 107;
Qy 5 KSTLLLAT---AAALIGSTVFGT-----VVGLAKVKYRGVNPQTQGVISQLGLIDSVAF 55
Db 820 EETILVVTLEKSQLATEEFYKSFDFEVPQWASSELNNKNDLEGVQ-----NTLVS 874
Qy 56 KPSIANFTSDYQSVKALLNGKTFDPKSSEPTDFVSKPDF-----LTNNGRTVLEIPKKY 110
Db 875 KRPIIPFVKDFD-----KSTLLVDAVKLEDFVKDYSVVHNGKI-----KF 915
Qy 111 QVVISSESPEDKRRFLGFLHKEKLEGN--IAQATKFIYLLPLDMPKALQ----- 163
Db 916 EVISWSSNPLSE-EQIR-----NFEQSNTEITHSGTATV-----RISIGENTNKW 959
Qy 164 ---YSYIVD-----KNFNLLIHH-----PL--SNFSAQSIKPLAL----- 193
Db 960 SDTTFVVEGFPSEKADNQKIVNDYVDLPETDTTKAKPILLARNVNTLAPRSLTSHF 1019
Qy 194 -----TRSSDFTAKLNQF--NNQDELWYLEK----- 218
Db 1020 EINTHRNRRRVNVIKSVEDVKNDQTKAIVLEASAGSGNERATKEYAYRFEGLNNLQ 1079
Qy 219 ---PFDLEALKANIRLOTAFSEKGNLVDPFVYSF--IRNPQNK-----EWASDL---NQ 267

Db 1080 EDIYNDIEAIARRIN-----AAQRANNIELISSFRLADPRSKKLASEALASDLFFANS 1133
Qy 268 DQKT-----VRLYLK-----TEPSQAKTIL----- 288
Db 1134 NSGTIFSANIDSSGAQDISKVRFTIGTRDGRSINSSTLVEATIFSPNAQLIINKAPEL 1193
Qy 289 -----KQYKYKDETFUSSIDLKASNGTSLFANENDLKQD 322
Db 1194 APFAGQVQKTNANAQTSDLRVIAPQONQEYEL--YKIVSLRAHNSTNRAIDGSEVSQ 1251
Qy 323 LDVLLDLDVSY-----EGG--QSETITSNQVKE-VPASERS-----LKDVKFKKQDQ 368
Db 1252 ISVNIKQSGFTPKVYTRIFSGYLSNDKFAKOKVDYIISGRSRANLVINAIKQKSVI 1311
Qy 369 KPRIEKFSLFYDALSFYQELVSKPN-----SIMDLVNATLARNLRS----- 414
Db 1312 ELEPSDFS-PDFQAGNLNLEIGTITTKTLNTAVVEVLVSAGNTTNSRYTKSYNVEIAG 1370
Qy 415 -----LGKYNFLPDDLASHLDYYFL-----VSKAKIKQSSITKKLFIEL 453
Db 1371 FKPEGAAENRIKLKCY---INDLANRKNPVLLQSVDKENTNASDLKASDFEIVKNAGFSG 1427
Qy 454 PIKISLXSSILGDOEPNI-----KTLFEKEVTFKLDNFRDVEIEKAFGLLYPG 501
Db 1428 DIIQTIESIARKDRNPEVALVTIKVVTGEGDKVAEDKYTIEIPGFASNERIRHILGLISAY 1487
Qy 502 VNEELEQARKAQRASFEKSKKGLKEFSQKBEENSKAINNQBEGLEEDDN----- 551
Db 1488 TSSMSEALYQRAKLTQNRKPNQVSLLAITQ-----DQLSVNSNFTIDQLSNHSQLHPSVLS 1543
Qy 552 ITERLPNSP-----IQOQENAGLQSPDKPMIHKDQVQNY-----YLAQSQJOEL 599
Db 1544 VPKENBETTAITIRIQAPTSNAGVAD-----KDIATFTYBESGYLTNGQVPNF 1595
Qy 600 IKAKDYTKLAKLSNRHTYNIIS-----RLKEQLFDVNPRIPSSRDIEKAKFVLDKT 651
Db 1596 LKAQDYAKKAENQRKPIILTDGDKNTPVRLTINSFTTQPTTEGLKLEITKVEPDPT 1655
Qy 652 EKN--KYWQIYSSASPVFQNKWSLFGYYRYLLGLDPK-----QTIHELVLKQKAGLQFEG 705
Db 1656 NNNNAKVTITVAGSGINENRSYTHEIERGFASEGKAVNEANVKEYIESDKA----- 1709
Qy 706 YENLPDFNLEDLKNRIKTPPLPSQKNFKLLDFNNYDGEIKAPE--FGLPLFLPKKE 763
Db 1710 ---LPS--LVVGIEKERVEAQSISSNLF-----IQQDPDKGLSLTIEKV 1750
Qy 764 LRRNSS-----NSGGSNSNSPWEQEIISQFKQNLNSQDOL----- 800
Db 1751 NVQOQSLIVEIRVSAGSQNLASQTYNHTITGATQEQTKNYALLRKVTLPGRRTPSLV 1810
Qy 801 -----AQFSTKIWEKIIIDENEFQDNRLQYKLLKDLQ--ESW-----INKT 840
Db 1811 EPOSRSVKTVSOLQTDNKIDQPGSGEY--SS1QLSIKEIKELERDKSWAVVIVEATINGL 1869
Qy 841 RDNLYWTYLG-----DKLKV-----KP--KNLEAKFRQI----- 868
Db 1870 KDTYQLSSGFATQAKKVNIDLQVKWYKEAQTKNPTLKEVETQKANIIVNLNVNDF 1929
Qy 869 ---SNLQELLTAFVTSAAALNNWNYODSGAKSTIIIEEIAELDPKVEKVGADYOLKF 925
Db 1930 IIPTNLENISIDLLSVS-----EKSDSETD---QKIALRVKVSAGSGSDYE-EF 1977
Qy 926 HYAIGFDNAGKFNQEVIRSSRTIYLTSGSKLE--ADTIDQLNOAVKNAPLQSGOSFY 983
Db 1978 YVY---DYEG-FAERRVQGSIRAI---TNVIANVENRAQVFLKVTSSINKHPSEITKED 2029
Qy 984 LDTERFGVFKLATSILAVQHKQEKTLPLKLNNDGYTLIHDKLKPKVPIQLSSSPKDWKF 1043
Db 2030 LDLPNSG-----VENIKVEITEVKKIENDATTL--DLNRVTAIVRVSIDTQ----- 2073
Qy 1044 EGKLNQNGSQNQVNV---STFGSIIESPFSYTFNQEDADLDQDQ-----QDDSRQGNNS 1094
Db 2074 --SITSHTATVESNIFGFTTRASYLIQQLINSNARPSVLQTKDGSQQVVSQEEKKILPS 2131

Db 2535 CKQLNSVI-----NVYCPYV-----LKKHNL-----YKSIDNIKLVDKVKYVLSNLYHA 2581
Qy 882 AALSNNMNYQDS-----CAKSTIIPEETAEADPKVKEKVGAD----- 919
Db 2582 FIFSKNIKYQHNFNIPPLRLLFLYDVPYKFNKDIKNNVRFKFFKYNLEKECIYIKY 2641
Qy 920 --VYOLK--FHYAGFDNAGKQNEVIRSSRTIYLKTSKSK-----LEAD 963
Db 2642 VFYILLKYIFLKYKYDKKNTIDIKNFSINSIPKLNLEKKYIYVLYREDIYYLE-N 2700
Qy 964 TIDQLNOAVKAPLG-----LQSFYLDTERFGVFKLATS LAV-----QHKQK----- 1006
Db 2701 NINQINVEQNHNFEHFHYDLLFQIIKPFNFVFLNIKILIRYILHEHLQNCIITNTF 2760
Qy 1007 -----EKLTPKLLNDGYTLIHDKLPKVPPOISS-----SPEKDWE 1044
Db 2761 QHSIFNITDVNSLKVVIQYKQKYNKD-----HSSQKQEDIFKGSQNELLTNEKTGTGE 2815
Qy 1045 GKLNQNGQS-----QNVNVTFTSGIIESPYFTSNFQEDADL----- 1080
Db 2816 KKQNLIHQDDHLKRNKRNKHVDEYNNYNNINEKDQ--QDPNKNTSNQNDYLVNTNQITIN 2873
Qy 1081 ----DODQDD--SRQGNLSLDQEAQLLKQKLAILLGN---QFIQYQOQNDKEIFEI 1130
Db 2874 SFGYDDDDNNSIYSNEYNSDFTENCFAYNKONINLAHLQKMPCKKYKKNILNSIKYF 2933
Qy 1131 I-----NVEKVSELSFRV-BFKL-----AKTLEDNGKTIR-----VLSDETMSLIIVNTT 1173
Db 2934 LENLLNNIQKVLKAFVFPYVYVILGCTSPKLLDRMKVKVQVQIKNSHTNHTVNSK 2993
Qy 1174 IE--KTPMSA-----VPEVFTKWVEQ-----YDPRTPLAAKTKFVLKF 1211
Db 2994 GDGRKNPDASQEVNRKENQIKPNI--DKKNVDNNDKVEKDNLLDDIKKCSLKNKKIIEY 3052
Qy 1212 KDQIPVDSGNSDKWLASIPLVTHQMLRSLPVVKTIRELGLKTEQOQQOQQOQQOQ--- 1268
Db 3053 LKK-----YSNIWYTDIE--KYHSKSISKML--IEKRLNDI--LRYRQNLVHQLNKKRIDK 3101
Qy 1269 ---QOPQKAVKBEELSTYNPKEFNI-----LNPLTKAHR 1302
Db 3102 ICAQFLRNNYIKFSNFYFCVDIDNIDVTDFELIILNRYLYFRTLEQINEFVNKPT 3161
Qy 1303 LTLNLSLVNDPNY-----KIEDLK-----VTKNBAGDHQAFSLRANNIK 1342
Db 3162 FYLQOLKEN--PSFVLPPIFVYSCIEIENLHNLHKSTKCVINR--EHLISF--GNKIK 3215
Qy 1343 RLMNTPITFADYNPPFYNE--DWRSIDKYLN---NKGNVSSHQOQAGGNGSLIQLN 1398
Db 3216 K-----YEKIOTNNIDDDNIMELIHLRNQYDILSHGCIILYNI PYNKNQLYFL 3265
Qy 1399 KNIPETFTPALIAL---KDRNNTN---LSNVSDKII---MIKPKYLVERSIGVPMWSTGL 1449
Db 3266 HNAKKKVVPHLIFIKHNHDKNTYELIINNNDKICITDNMNVQNVNHNINSFNDKDN 3325
Qy 1450 DGYIGSEBTKD---GTSSS---SOQGFQDDF---IQALGLKNTEYHGKLGUSIRIFDPGN 1501
Db 3326 IHMINDEYKKKYINTSFTLLSKKNKEDSYCNLLQFISIIIEQSFQEKYKNI--IYIEYS 3384
Qy 1502 ELAKTKDASNKGBEKLKSLYDLFGNYLNEYKSKPKTAKGWTIHDPQKYPNPQKLP 1561
Db 3385 NLWKNMKKITDIEBK-IEKSYQKNRY--KYNNK--PAYFKGY-NIY-----EKANYEIQ 3435
Qy 1562 ENYLNVLNLQWPKVTL-----YNSSDFITNLF-----VEPEGSDRSGT 1600
Db 3436 FNNYCLVTYKKQILLKCNLYDNYTICHNSFVLVNSYDDVISFEKNPTEYINVDSSFK 3495
Qy 1601 KKKQVIOKQVNNVADWGSAYLTFWYDKNITNOP-----NVITANIADVFIKDVKEL- 1653
Db 3496 KLEQI-----NNFT-----YNNNNNNNPIENEGFCLVSLNDEGQFIGKDXSCC 3540
Qy 1654 --EDNTKLIANIITQWFWNISGSKEKFPYKPTVF-----FGNWNENSSNNSQAQTPTW--- 1704
Db 3541 IKYDNKYVIFQNKQKQOTFINNYNDYLNQNIIFDILNLQKNKN--DLNLTILYLKTYLVEI 3599

Qy 1705 -----EKIRGFALQ-----ALKSSPDQKTRTPVLVTNA 1733
Db 3600 LCHALVELGHRPIYPLGLTIQLSAIKFLALFYKENKLFISPLRKTYEQNYQTFLTDCDI 3659
Qy 1734 PLPLMKVGLPGFQNGPNFKTQ-----DWRLVFQNDNDNQIAALRVQEQDRPEKSSSDKD 1786
Db 3660 PIHQKL-KQKPKNSNKNQNTINTPVNWN--TKRDIKHYHKYINLYDKILYSOTITNKT 3716
Qy 1787 KQKWIF 1793
Db 3717 KQKIKY 3723

RESULT 32
T13030
microtubule binding protein D-CLIP-190 - fruit fly (Drosophila melanogaster)
C;Species: Drosophila melanogaster
C;Date: 13-Aug-1999 #sequence_revision 13-Aug-1999 #text_change 17-Nov-2000
C;Accession: T13030
R;Lantz, V.A.; Miller, K.G.
J. Cell Biol. 140, 897-910, 1998
A;Title: A class VI unconventional myosin is associated with a homologue of a microtubule
A;Reference number: Z17588; MUID:98139549; PMID:9472041
A;Accession: T13030
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: mRNA
A;Residues: 1-1690 <LAN>
A;Cross-references: EMBL:AF041382; NID:g2773362; PID:g2773363; PIDN:AAB96783.1
A;Experimental source: strain Oregon R
C;Genetics:
A;Cross-references: FlyBase:FBgn0020503
C;Keywords: cytoskeleton

Query Match 2.88; Score 276.5; DB 2; Length 1690;
Best Local Similarity 20.08; Pred. No. 0.00065;
Matches 270; Conservative 233; Mismatches 458; Indels 387; Gaps 61;
Qy 343 SNSQKEV---PASERSLKDVRVKFQKQDKQKPRIEKFSLYBYDALSFYSQQLVSKPNS- 398
Db 352 SSTPVKPILATPKSQFSQMLLREKQHQHVKLMVERDLREDQAQNALQKKNINELKAR 411
Qy 399 IKOLVNA-----TLARNLRFSLGKYNFLPDDLAHLDYFLVSKAKIK--QSSITKKL-- 449
Db 412 IVELESALGNRKKTEBLOCSIDEAQFCGDELNAQSQVY---KEKIHLESKITKLVA 467
Qy 450 -----FIELPKISLK---SSILGDO-----EPNIKTLFE 476
Db 468 TPSLQSLTPDPLPSDDGALQEEIAQLQERMTIQOKEVESRIAEQLEEEQRLRENVKYLNE 527
Qy 477 KEVTFKLDNFRDVEIEKAFGLLYPGV-----NEELEQAKQAQASFEKEKSKG 525
Db 528 QIATLQSELVSKDEALEKFSLSGEGIENLRRLALLKEENEKQAEQAQ--AEFTRKLAES 586
Qy 526 LKFSQOKE--ENSKA-----INNQGLE-----EDDNITE---RLPENSPIQY 564
Db 587 VEVLRSSELQNLKATSDLSLESERVNKSDECIQLQTEVRMRDEQIRELNQOLDEVT--TQL 645
Qy 565 QQENAGLAGSPDPYMIKVQNOYRYLAKSQIOELIKAKDYTKLAKLLSNRHTYINISLR 624
Db 646 NVQVADSSALDDMLRLQKEGTEKSTLLEKTELVOIKE--QAAKTLQDKE-----QL 697
Qy 625 KEQLFDVNPRIIPSSRDIEKAFVLDKTEKNKYMQIVSSASPV-----FQNKW 671
Db 698 EKQISDL-----KQLAEQEKLVREKTE--NAINQIQLEKESIEQQLAKQNELEDFQKK 750
Qy 672 SLFGYRYLLGLDPPKQTIHELVLKQ---KAGLQFE-----GYENLPSPDFNLEDLKNIRIK 724
Db 751 SESEVHLQEIKAQNTQKDELVESGESLKKLQOQLEKTLGHEKQLA--ALEELK--KEK 806
Qy 725 TPLFSQKDNFKLSLLDPNNYVDGEIKAPFGLPLFLPKELRRNSNGSGGNSNSPWEQE 784
Db 807 ETIIEKEQEQLQLOSKSASESALKVVQVL-----EQLOQQAASGEGSKTVAKLHD 861

Db 1230 EFRGYKTHRLDGNIGLOE--RKIIIDQFNNNVEYKDEKQKQPCNCPGNNMNSGNENM 1287
Qy 1053 -----SONVNVSTFGSIIIESPYSTNPQEDADLDQDQDD--SQGNNSLD---NOBAGLJK 1104
Db 1288 NMSVNNMNSVNNENMNSGNENMNSGNENMNSGNENMNSGNENMNSGNENMNSGNENMNSG 1347
Qy 1105 OKLAILLGNQFOY--QONDKEIEFLINVEKUSLSFRV--EFLAKLTLEDG---KT 1157
Db 1348 NENMNSGNENIKWISSQNEKOTSQSV---KISLKEBEINDFQIMDDKNVNGNQDAM 1404
Qy 1158 IRVLSDETMSLIWNITTEKTPMSAVPEVFTKWVEQYDPRTPPLAAK-----TKFVLK 1210
Db 1405 IFILSTRSGSLNLTQDT---VLIIDS-----DFNPHQDIOAWCHRRIGQKVVVK 1454
Qy 1211 PKQDIPVDSGNSIDKWLASIPLVIIHQMLRLSPVVKTIREGLKTEKQOQOQOQOQOQO 1270
Db 1455 VPRPITLSGVEELVFK-KAQHKLISNDKVIQAGLFNK-----IYNDEDQNKLDIIOR 1507
Qy 1271 POKKAV-----RKEELETYNPKDE-----FNILNPLTKAHLRLTSLN 1308
Db 1508 NQKNDMTHTPTNPLLLNYMKRNEELEVFLDPDKRYFGEQYFSLN-----TLNVEN 1561
Qy 1309 VNNDPNYKIEDLK-----VKNKAGDQO---LAFSLRANNIKRLMNTPIITFADYNP 1356
Db 1562 DSGQFTYMSDEKEENETVLSIIKKEKEEGEDDEENQORDNKEDQDED-----1613
Qy 1357 PFYNNEDWRSIDKYLANNKGNVSHQOQAGNGSGSLI---QRLNKNIKPETFTPALIALK 1414
Db 1614 ----KDDDKDKDKDKKEEBEKKRKHILNANNNGIQNGSSINEGVKEK-----ILDEV 1664
Qy 1415 DRNNTLSNYSKLIIMKPKY-----LYERSIGVWSTGLDYGISEQTKOQTSSSSQ 1468
Db 1665 CNNTKCVKVSNERLIFPKRKHDTDDIQCDEKIKENEEDVDNIIQNKNN-RLRWECOK 1723
Qy 1469 KGFDDQFIQALGLKNTYEH---GKGLSIRIFDPCNELAK-----IKDASNKKGBEKL 1519
Db 1724 DKDDDI-----NSNIHWDEKKIYMSSEKDDTKKEYSDTHDPYINDKQVQDEEDYY 1776
Qy 1520 KSYDLFPKNYLNVEK---KSPKIAK-----GWTNIHPDQKEYPNPNQKLPENYLN 1566
Db 1777 -GFILKEENQDIEKILIKSNKLINKDELPAFLFYDDTNDSPDK-----INLRSRKVIN 1830
Qy 1567 LVNLNQPKVT-----LYNSSDFITNLVPEPEGSDRGSGTKLQVIOQVNNYADWGA 1620
Db 1831 INLMQEEKLTEKQPLKIDSS--PNLLSSVE---KDLGRNKDIOVKSOMEHNN---DIT 1882
Qy 1621 YLTFWYDKNIIITQPNVITANIADVFKDVKLEDNTKLIAPNI---TQWMP-NISGSKEK 1677
Db 1883 TLEVKDREBIKEHLETTQNISSLIND---LEINKTLTNEVHVSTKSPYNMRSSKR 1939

RESULT 35
E71619
RAD2 endonuclease PF0265C - malaria parasite (Plasmodium falciparum)
C;Species: Plasmodium falciparum
C;Date: 13-Nov-1998 #sequence_revision 13-Nov-1998 #text_change 09-Jul-2004
C;Accession: E71619
R;Gardner, M.J.; Tettelin, H.; Carucci, D.J.; Cummings, L.M.; Aravind, L.; Koonin, E.V.;
Perteaux, M.; Salzberg, S.; Zhou, L.; Sutton, G.G.; Clayton, R.; White, O.; Smith, H.O.
Science 282, 1126-1132, 1998
A;Title: Chromosome 2 sequence of the human malaria parasite Plasmodium falciparum.
A;Reference number: A71600; MUID:99021743; PMID:9804551
A;Accession: E71619
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-1516 <GAR>
A;Cross-references: UNIPROT:O96154; GB:AE001383; NID:G3845135; PIDN:AACT164
A;Experimental source: clone 3D7
C;Genetics:
A;Gene: PF0265C

Query Match 2.8%; Score 272; DB 2; Length 1516;
Best Local Similarity 18.6%; Pred. No. 0.00086;

Matches 317; Conservative 254; Mismatches 543; Indels 594; Gaps 82;
Qy 36 GVNPTQGVISGLGI-----DSVAFKPSIANFTSDYQSVKKALLANGKTFDPKS-----83
Db 2 GYGLWSIVSPVGRVNPFIPTGRIAIADVSIWLYELTYANNVKDLRN-KSPDNNSIFND 60
Qy 84 --SEFTDPSKFDPLTNNGRV-----LETPKKY-----QVVISSEFPDDKFRFLQ 130
Db 61 LWIDSENISS-EIKTDNIIKAHLFFFLRICKLYNIRPIFIDGNPPPELKRKTIPOR 119
Qy 131 HLKLEKLEGNIAQSAKTIY-----KNNQDELWVYLEKFPDLKALKANIRLOTADFSFEKGNIV 244
Db 120 NIKKRYEKKFKAELVYNYQRTLLNSMKSNNKQNDNSNIEDKTNPNTKNTQNS 179
Qy 151 -----LLPLDMPKALGO-YSYI---YDKNNENLIIHPLSNFSAGSIKPLALTRSS 197
Db 180 NTQNSKNTPNKINADISKSSLIQIYDDIKEKDSLSNLSVEH-VGNVPV-SVKDV-LTICN 236
Qy 198 DFIAXLQ-----FNNQDELWVYLEKFPDLKALKANIRLOTADFSFEKGNIV 244
Db 237 DDLSKIKKIEMITDFGPVLFLGEQDGMGTVENIKLD-----NRNKDNNLS 285
Qy 245 DPFVYSFIRNPONQKEMASDLNQDKTVRLYLRTFESPOAKTILKDYKYKDDTFESSI---302
Db 286 YSINTYKVDVNNND--DDKDKKENINEVRDQ-----KNYYVYKNNENINNIYL 334
Qy 303 -----DLKASGTSIFANENDLQOL-----DVL 328
Db 335 DDDDEKEDIQNGV---YNNDDIDEDQIRIKHMARKKYYESIPKTFKGLCMWRPVDII 391
Qy 329 DVSDFVGGQSETITSNOKVPVPASERSLKORVFKKQOKPRIBKFSLYEYDALSIFYQ 388
Db 392 DISNY-----NTEMLEI---SETLKVHENKFK-----415
Qy 389 LQEL-VSKPNSIKDLVNATLARNLRFSLGKYNFLFDDLASHLDYDFLVSKAKIKOSSITK 447
Db 416 -QHLNVLDENNETPVVNNLLKNIN---KKN---DDLIEGGE-----K 452
Qy 448 KLFIPLPIKISLSSILGDOEPENIKTLPEKVTFLKDNFRDVEIEKAFGLLYPGVNEELR 507
Db 453 KSFINL---INVDS-----CYSSSSNRLE 473
Qy 508 QARKQARAFE---KESSKGLKEFSQOKBENSKAINNQOEGLEEDDNITERLPENSPIQ 563
Db 474 NDENIERKINPITNDEKISININNNNNNNNNNN--NDNNNDNDVIEHNKNNMIY 532
Qy 564 YQENAGLAGSPDKPYMIKQVQNYRLAKSOIQELIKAKDYTKLAKLSNRHTYINSLR 623
Db 533 -----DNKNYVE-----CSSHAKINDNGISNKNINILELPNNLDTSIFLE 572
Qy 624 LKEQ--LPDYNP---RIPSSRDIEKAKFVLDKTEKNKYWQIYSSASPVFNKWSLFGYVR 678
Db 573 GKDEYKVVYVNNKEEIRIPLFKEINKE--IPFKLPLKLYOILQD-----IKBEWYTNRIK 626
Qy 679 YLLGLDPKQTIHELKVLGQKAGLOPRGY-ENLPSDFNLEDLK-----NIR-IKTPLFSQK 731
Db 627 AKSKDDMDVFSQ-----VQLETYVYRMKTDPEIEKLIKMAENIQSVGELLINK 677
Qy 732 DMPK-----LSLLDFN-----NYVDGEIKAFEGFLPL- 758
Db 678 DLSKNTDINIKDYNVLQKKKKKKKKFLNDILTNTFTTESKYQDLYVKGESKEDIK 737
Qy 759 -----FLPKELRRNS-----SNSGGSQNSNSPWEQEIISQPKDNLN 796
Db 738 NQIDFVTOCYRNNDIIRDTDKSDIFKNIRKIDNNKYEIYNLELEQEEINEKKNYKN 797
Qy 797 QDQLAOFSTKIMEKIIDGENEF-----DQNNRLOYLKLLKLOESWINKTRDNL-----844
Db 798 NDSKTFFLKI-----ENEFKKDLLDDSQIFGDSLLADIKE--YNYTADNLNNDNEN 848
Qy 845 -----YWT-----YLG-----DKLVKPKONLEAK-- 864
Db 849 KSLYEDGENPITRNEPITNBEYBKNNIIVISDQKYNEEDIIFKDKIKKEKKNDDSSD 908

QY	865	FRQISNLQELLTAFTYSAALSNNWNYQD--SGAKSTIIIFEEIAELDPKVKRGADYVQL	923
DB	909	FENCSS-VQEKI---YANEKI BEYNNKNDKSSSSSSIIIEEIKYKKEKDELVSPNLCVL	964
QY	924	--KHYAIGFDDNAGKNQSVIRSSRTIYIKTSGSKLEADTTIDQLNQAQKAPLGLOS	981
DB	965	LDEFEHSDNLNN-----YISVSSD-----DMKTNYSKNITGVKE	1000
QY	982	FYLDTREGVFQK-----LATS LAVQHKKQEKTLPKKLN--NDGYTLIHDKLKPKVIPQ	1033
DB	1001	NKVDTKNVEYDKGDGVIETSPEDSHKLESKPDNNNIYDND-----DELEKNLSKD	1054
QY	1034	ISSPEKDWFGKLN-QNGQSQNVNVSTFGSIISSPYFSNFQEDALDQGDQDSDRQGN	1092
DB	1055	YISDVDKNHNVIYNIERGEEREN-----EFVENKIQT-----ESHKSN	1095
QY	1093	NSLDNQEAGLLKOKLA-----ILLGNQFIQYQONDKEIEFEIINVEKVSLSFRVE	1144
DB	1096	EFICTENKSLRKQYMSKEDISNVRILKSDDINNLSKQN-----YFEIILLDKKQVMDNFQMN	1151
QY	1145	F-----KLAKTLEDNGKTI RVLSDETWSLI VNTTIIKTPEMSAVPEVFTKWEQYDPT	1199
DB	1152	IEQNNDKLKEDKLDEGAFVEYLED---NKIIDSYIKETNKE-----NEELIKEY----	1197
QY	1200	PLAAKTKVFLFKPOI PVDGSGNISDKWLASI-----PLVIHQMLRLS-----PV	1245
DB	1198	---KKLK-----KNNIETDNEMDDI KLLNFFGIPYIQSPCEAAQCSYLNKNKYCDAI	1249
QY	1246	VKTIRELGL---KTEQQQQQQQQQQQPOKQKAVRKEELETYNPKDEFNI-----LNPL	1297
DB	1250	ISDSDVLVFGSKTVIKNFNKKTVVEYKAL--BEKGLGYQ-ESLIINISLLCGDYT	1306
QY	1298	TKAHLRLTSLN---LVNNDPNYKIEDLVKINEAGD--HQLAFSLRANNIKRLMNT-----	1347
DB	1307	IGVHGIGIVNALEIKAPPNF--EDLKILKDIVSNPFKIDKQWYNEEIQQFLNTHKNYK	1364
QY	1348	-----PITADYNPPFYNNEDWRSIDKYLNKGNVSSHQOQAAGG	1387
DB	1365	LNWIFPNFPDREYVKFKYPKVCTDIKKFEHWPDIKSITKFLHKTNTISEKVL-----	1420
QY	1388	NOGSLGIQRLNKNI-----KPETFTPAL 1410	
DB	1421	NVLNPILOKYNVNVRTYQSKIEDFFPL 1448	

RESULT 36

152300

giantin - human

N/Alternate names: gcp372

C/Species: Homo sapiens (man)

C/Date: 01-Nov-1996 #sequence_revision 01-Nov-1996 #text_change 09-Jul-2004

C/Accession: I52300

R/Sohda, M.; Iizumi, Y.; Fujiwara, T.; Nishioka, M.; Ikehara, Y.

Biochem. Biophys. Res. Commun. 205, 1399-1408, 1994

A/Title: Molecular cloning and sequence analysis of a human 372-kDa protein localized in

A/Reference number: I52300; MUID:95100974; PMID:7802676

A/Accession: I52300

A/Status: preliminary; translated from GB/EMBL/DBJ

A/Molecule type: mRNA

A/Residues: 1-3225 <RES>

A/Cross-references: UNIPROT:Q14789; GB:D25542; NID:G662389; PIDN:BAA05025.1; PID:9808865

C/Superfamily: giantin

QY	106	IPKYOYVVISFEPSDDKERFLGFHLKLEKEDGNIAQSATKFIYLLPLDMPKPAALGOYS	165
DB	1704	LSKKPQSLMSKOSLSSEVQ-----DLKHQIE-GNVSKQAN-----LEATEKHQNTN	1750
QY	166	Y-----IVDKNFNNLIHPLSNFSAQSIKPLALTRSSDFIAKLNQFNQNDQLWV	214

Query Match 2.8%; Score 271; DB 2; Length 3225;

Best Local Similarity 18.8%; Pred. No. 0.0028;

Matches 308; Conservative 246; Mismatches 507; Indels 578; Gaps 76;

Db 2706 MSLSQNSRDHANEELDELKRRKYDASLKELAQKEQL---LNRE-----RDALLSETAFS 2757
Qy 1034 ISSSPKDWFE-GKLNQNGOSQNVNVTFGSIIESPY-----FS---TNFQSDAD----- 1079
Db 2758 MNSTEENSLSHLEKLNQQLSSQDEQLHLSSQLEDSTYNQVQSFSAWASLQNERDHLWNE 2817
Qy 1080 -----LDQGDQDSS-----RQGNNSLDNQEAGILLKQKLAILLGNQFOYQY 1120
Db 2818 LEXFRKSEEGKQSAAPSTPAEVOSLKKAMSSLQNRDRLLKE-----LKNLOOQYLO 2872
Qy 1121 QN-----DKETEFEIINVE-----KVSELSF 1141
Db 2873 INQETELHPLKALQOYQDKTKAFQIMQBELRQENLSHQHEDQLRMEKSSWEIHERM 2932
Qy 1142 RVEFKLAKTLEDNG-----KTIRVLSDETMSLIVNTTIEKTPMSAVPE-----VFD 1188
Db 2933 KEQYLAISDKDQQLSHLQNLRELASSSSQTPKLVQVQROQASPSASPDGSQLVYE 2992
Qy 1189 TKWVEQYDPTPLAAKTKFVLKPK---DQIPVDGSGNISDKWLASIPLVHQQMLRLSPV 1245
Db 2993 TELL-----RTQJNDSLKETHOKELRIQQLNSFNFSQELLEKNLTLQLCOTQSOLR----- 3043
Qy 1246 VKTIRELGLKTEQQ-----OQOQOQOQOOP-----QKXAV-RKBE 1281
Db 3044 -----ENQHYGDLNHCVALEKQVQELQAGPLNIDVAPGAPQEKGVHRKSDP 3092
Qy 1282 LETYNPRDEFNPLTKAHLRLTSLNVLNNDPNYKIEDLVKVNKEAGDQOL-----AFSLR 1337
Db 3093 BELREPOQSFEAQ-----QQLCNT--RQEVNELRLKLEERDQORVAENALSV 3140
Qy 1338 ANNIKRL-----MNTPI 1349
Db 3141 EQQIRRLHSEWDSRTPI 3159

RESULT 37
T30822
lmp1 protein - Mycoplasma hominis
C:Species: Mycoplasma hominis
C:Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 09-Jul-2004
C:Accession: T30822
R:Jensen, L.T.; Ladefoged, S.; Birkelund, S.; Christiansen, G.
Infect. Immun. 63, 3336-3347, 1995
A:Title: Selection of Mycoplasma hominis PG21 deletion mutants by cultivation in the pre
A:Reference number: Z18884; MUID:95369882; PMID:7543881
A:Accession: T30822
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1365 <JEN>
A:Cross-references: UNIPROT:Q49525; EMBL:U21962; NID:G790243; PID:G790244; PIDN:AAA81013
C:Genetics:
A:Gene: lmp1
A:Genetic code: SGC3

Query Match 2.8%; Score 270.5; DB 2; Length 1365;
Best Local Similarity 20.1%; Pred. No. 0.00086;
Matches 328; Conservative 235; Mismatches 565; Indels 501; Gaps 80;

Qy 283 QAKTILKDYKYKDTFLSSIDLKASNGTSLFANENDLKQDLVDLLDVSDFYFGQSETIT 342
Db 41 QENLLQNFKKLQQLNELLYGKIVNINVFHEQVLOGSL-----KIN 84
Qy 343 SNSQVKPVPASERSLKDVRVFKDQKQPKRIEKFSLYEYDALSFYSQQLVSKPNSIKOL 402
Db 85 NKSETKAIESETLKDAILTISKIKNQINQEL-EF-----AKFNEIKDK 130
Qy 403 VNATLARNLRPSLGYNPLFDDLAHLDDYFLVS--KAKIKQSSITKKLFIPLPIKSLK 460
Db 131 LQEVYKNEK--SKQEYEHKQNIENELNKTPTISLESTLIEIQNATNL-----IKLNE 183
Qy 461 SSILGDQEPNKTILFEK-----VTFKLDNFRDVEIEKAFGLLYPGVNELEQ 508
Db 184 ST---KSKONIDNLNAKEQLKASISQANQLLPQLSDN--DSEIATA-----KKSIDAEIKN 234

Qy 509 ARKA-----QRASPEKEKSK-----KGAKESQOKE-----ENSKAINNOBGEEDNI 552
Db 235 ANQAVASNTASQMSQASSLDAKVAIITKQLETFNFKDKEAFNELKQTRNQ--IQEFINT 292
Qy 553 TERLPENSPIOYQOENAGLGASPDKPYMIKQVQORYYLAQSIOQ-----ELIKAKDYTKL 608
Db 293 NKQNPYSELISQ-----LTSKSDSKNSVTBSSN-----KSDIESANTELAQ-----L 336
Qy 609 AKLLSNR-HTYNISURLKEQLFDVNPRIPS9-----RDIEKAFVLDKTEKNKY 656
Db 337 AKANADKVQADNLAKSIKEQL-----NNSVSNANTLSAKLTDKONTTIOAKTELEKEVQKAD 393
Qy 657 WQIYSSASPVFQNKWSLFGYRYLLGLDPKQTIHELKVLGOKAGLOFEG-YENLPSPDFN- 714
Db 394 QAIKSNNTASMSQAS-----SLDAK--VABITKLETFNKDKKEAFNELKQTRNQ 442
Qy 715 LEDLKNIRIKTPLFSQ-----KDNFKLSLDFNNYDGEIKAPFGLPLF----- 759
Db 443 IOEFINTKNPNYSELISQLTSKRDSKNSVTSSNKSQSDIESANTELKQALAKANADKVQ 502
Qy 760 -----LPKELRNNSGSGGNSNSPWEQELISQFKQONLSNQDQ-LAQFSKTIWEKIIGDE 815
Db 503 ADNLAKSIKEQLNNSVSNANTLS-----AKLTDKONTTIOAKTELEKEVQKAD 550
Qy 816 NEFDQNNRLOYKLLKDLQESWINKTRDNLVYLGDKLVKPKNNLEAKFROISNLOELL 875
Db 551 QAIKSNNTASMSQASSLDAKVAIITKLE-TFNKDK-----EAFNELKQTRNQI 600
Qy 876 TAPYTSAAALSNNWYQ-----DSGAKSTIIFEETAEALDPVKKE---KV 916
Db 601 QEFINTN--KNPNYSELISQLTSKRDSKNSVTSSNKS-----DIESANTELKQALAKA 653
Qy 917 GADVYQLKPHYAIGFDDNAGKFNQEVIR---SSRTIYILKTS-----GSKLEADTI 965
Db 654 NADKVQA-----DNLAKSIKEQLNNSVSNANTLSAKLTDKONTTIOAKTELEKE-I 703
Qy 966 DQNLQAVK-NAPLQLOS--FYLDTERFGVFQKAT-----SLAVQHKKQKETLPK 1012
Db 704 QKANOAKSNNTASMSQASSLDAKVAIITKLETFNFKDKEAFNELKQTRNQIQEFINT 763
Qy 1013 KLNNDGYTLIHDKL-----KKPVIPIQISS9-----P 1038
Db 764 NKNNPNYSELISQLTSKRDSKNSVTSSNKSQSDIESANTELKQALNTAKAKSSIDNELRP 823
Qy 1039 EKDWFEGLKQNGQSNQNVNVTFGSIIESPYSTNFOED---ADLDQGDQDSDRQGNNSL 1095
Db 824 LKNDLQSKIEEFGPIRNTNFWISSKLETT--KNKLABELTKADAIKNNPSSSKQALKDS 881
Qy 1096 DNOEAGLLKOKLAILLGNQFOIYQQNDKEIEFEIINVE-KVSELSFRVEFKLAKTLEDN 1154
Db 882 SQO-----VQK-----LGNELL-----KTIIEFGKVKETKNSNIGYRL-FKLAQAEQFN 924
Qy 1155 GKTIRVLSDETMSLIVNTTIEKTPMSAVP9VFDTKWVEQYDPTPLAAKTKFVLKFKDQ 1214
Db 925 -----NSVDVK-----LKNWEEK---QTLSSKKQKL----- 948
Qy 1215 IPVDSGNTSKWLASIPLVHQQMLRLSPVVKTTIRELGLKTEQOQOQOQOQOQOPOKK 1274
Db 949 -----GNQSTK-----DYLQTLST8MT-----QESTIKK 973
Qy 1275 AVRKEELETYNPKDEFNINPLTKAHLRLTSLNVLN-----DPNYKIEDLVKIKNEAGD 1329
Db 974 VI-----VNIQAHIRNNLSQYRLKADKLIANMKRGYGDVIGIESLQKQDMLDD 1023
Qy 1330 HOLAF--SLRANNIKLMTPTIFADY--NP-----PFYVNEWMRSIDKLLNKNQNVSSHQ 1382
Db 1024 SVLSVDDSLKDDFNKALR---VLVDYTKNPPVSSWFFINKNFSIENYQNLRLILVREN 1080
Qy 1383 QAAGGNGQSGLIQRLKNIKPKETFTFPALIALKDRNNNTSNYSYSDKIMIKPKYLVRSIG 1442
Db 1081 EIL-LDKANDLDKRAETIK---FV-----DENINSLDQRAKRLKQELINAKN-- 1124

QY 1443 VPMSTGLDYGIGSEGTGDTGSSSQKGFDDQFIOALGLKNTYHGKLGSLRIFDPGNE 1502
DB 1125 -----DLGNFTLNHQKQFTAKDITPK-----ISLENKLEINQYLLPIIK-----EKA 1169
QY 1503 LAKIXD-ASNKKEGKELKSLDYDLFNKYNLEYEKSPKIAKGTWNTHPDQKEYPNQKLP 1561
DB 1170 VSKISEIEKNKKEBLEDIIRS-----NFYLWEKVEINKYISELNTKNQVELRS----- 1215
QY 1562 ENYLNLVLNQPMKVTLYNSDPITNLFVEPEGSDRGSGTKLKQVTKQVNNYADWGSAY 1621
DB 1216 -----NINENKWE-----SIKOTLDNLNIKENV-----LKEVI-----INNSNAQYSI-- 1256
QY 1622 LTFWDKNIITNQPNVITANIADVPKDVKELEDNTKLIAPNITOMWPNISGSKEKFKYP 1681
DB 1257 -----NRILSTVPE-----FIKVAQITRSN-----NLSRLAEIQPK 1288
QY 1682 TVP-----FGWENENSSMNSAQPTWTKIR-----EGPAL 1713
DB 1289 TLLDIDKNLKEVKKTLDNKTLNSDNIAKNEKIRLLVDRSELKESLURWFRQNEISD 1348
QY 1714 QALKSSFDQ 1722
DB 1349 QRIKNKLD 1357

RESULT 38
F90608
ABC transporter permease protein [imported] - Mycoplasma pulmonis (strain UAB CTIP)
C:Species: Mycoplasma pulmonis
C:Date: 24-May-2001 #sequence_revision 24-May-2001 #text_change 09-Jul-2004
C:Accession: F90608
R:Chambaud, I.; Heilig, R.; Ferris, S.; Barbe, V.; Samson, D.; Galieson, F.; Moszer, I.;
Nucleic Acids Res. 29, 2145-2153, 2001
A:Title: The complete genome sequence of the murine respiratory pathogen Mycoplasma pul
A:Reference number: A99512; MUID:21267165; PMID:11333084
A:Accession: F90608
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-2599 <KUR>
A:Cross-references: UNIPROT:Q98PE9; GB:AL445566; PID:g14090189; PIDN:CAC13947.1; GSPDB:C
A:Experimental source: strain UAB CTIP
A:Genetics:
A:Gene: MYPV_7740
A:Genetic code: SGC3

Query Match 2.8%; Score 270.5; DB 2; Length 2599;
Best Local Similarity 19.7%; Pred No. 0.0021;
Matches 384; Conservative 269; Mismatches 689; Indels 607; Gaps 95;

QY 39 PTQGVISQLG-----LIDSV--AFKP-----SIANFTSDYQSVKKALLNGKTFD 80
DB 699 PRSDVIKEINNQGTFELFDSIYBAFKRSLVTRGFISKANLDMYGPFKALYNNKPHS 758
QY 81 PKSS-----EFTDFVSKFDLTNGRTVLPIPKYQVIVSEFSPEDDKERFLGF 130
DB 759 VLSSGSINSNIYFKMFFDWISEIN--KENGNTF-----FNDVILDFTKE----- 800
QY 131 HLKKELDGN-IAQSATKFIY--LLPLDMRAALQYSVIVDKNPNLLIHPLSNFSQAS 187
DB 801 -LKQIKGGGATTOQKQDFIVREFKFKDKIFGSLSTTKLFSALFLNLEVVDIITYIKDP 859
QY 188 IKPL-----ALTRSSDFIAKLN-----QFNQDELVYLEKFPDLEALKANIRLQT 233
DB 860 VKFIEGPEVIVKIDFTKATSEIGLWYEQNKQKQDNKNYILLSEIDVLLSI-----LDS 914
QY 234 AD---FSEKGNLVDPPFVYSTRPNQKQWASDL-----NODK----- 270
DB 915 ADQNKVRLGMKIVSEIDFNKFLNPEEQDSIYSRFFKIHGENQDVKNFNNKLDGSKGQE 974
QY 271 -----TVRLYLRTEFS--POAKTILDKYKDETFELSSIDLKAS 307
DB 975 KYSNVSKGLVELISFLDIKSFVARLKQNSEFKLLPKIPNNPKYKTNIS--ISSKNLWAS 1031

QY 308 NGTSLFANENDLKQDLQDVLDDVSDYFGGSEITTSQVKPVPASERSLKDRVKFKKQD 367
DB 1032 FLDSIFHGQSD-----ENGVLGINNDAKISILVEMFNISDAGEFLKIS 1076
QY 368 QKPRI-----EKFSIYEYDALS-----FYBQLQ-----ELYSKPNISKD--LVN 404
DB 1077 DKPLYIPSPFKDDKKVGLGDLISVLQSLPOLFAGQVDKSKVEFTQLEILNQLDKNLSLQN 1136
QY 405 ATLARNLRFSLGKY-----NFLFDD---LASHLDYYFLV-----SKAKIQSSITK 447
DB 1137 RNLSQVELLKKVLPLEQDFENFEDKNVIEHDANVLLTNQFRFNSNSSLVNFSITAK 1196
QY 448 KLFIELEPIK---ISLKSIIIGDQEPNITLFE-----KEVTFKLDNFRDVEIKAFGLL 498
DB 1197 RL-LDAPSNHEAVNLASKQLKDSLGSVYEFELSSGSLKHPDYKLG-----KKAISLL 1248
QY 499 YPGVNELEQARKAQRASFEKESKGLKSPSOQKEENSKA---INNQEGL--BEDDNI- 552
DB 1249 ALFI-----RLRKESLSIENYQNDYKKIFDFINTKSDDENAFVNVVNSKTLTKDQVDDNFD 1303
QY 553 -----TERLPENSPIOYQOENAGLIGASDPKPYMIKDVQNKORYLAKSIOBELI 600
DB 1304 IAGFISGKGLASTSRL--ANELFEKEAEN-----YKNKEIQFELDSLNEARKLIE 1352
QY 601 KAKDY-TKLAKLNSRHTYINISRLKEQLFDVNPRISSRDIEKAKFVLDKTEKNKYWQI 659
DB 1353 ENKEYFIQIAAILLSPEKEYH-----PTLLNI-----FEYEDIHNNPFK- 1391
QY 660 YSSASVPQNKWSLFGYYR-----YLLGL-----DPKQTIHELKVLKQKAGLQ 702
DB 1392 YSITPNPFENKIALNELNKQASAINOSKIYISLGLSYALVDPMLNI-----LLPOVALW 1446
QY 703 FEGENLPSDFNLEDLKN-----IRIKTFLFSQKONFKLSLD-----FNNYVD 746
DB 1447 FTSLNTPENSONKDNKSNLAYLITNKVNLGRKNFEQIQNFVSFVLGNSNEGFFNN-FE 1505
QY 747 GEIKAPFGLPLFLPKELRRNSNGSGNSNSPWEQEI-----ISQFKD----- 791
DB 1506 NDAKS-----NLFVDLDYLELIGINENSRKKID--EQELVFGINFKKTIYEFIDAVTYPH 1558
QY 792 --QNLNQDQLAQFSTKIWEKIIGDEN-----EFDONNRLOYK--LLKDLQESNI----- 837
DB 1559 TLDNVVFNVDVASYLVKVNQAYLVNKNKTIYTGPIPTNNKLFNFKKDNKDESNLVNNG 1618
QY 838 -----NKTRDNLVYTLGDKLVKPKKNLEAK-----FROISN 870
DB 1619 SQFIIVGTDITSDVLY-----PVDENNQVNTSQQALVYVQKGFBRVKTFSFSPN 1670
QY 871 LQEL-----LTFYVTSAAALSNNWYVQDSGAKSTIIFEEIAELDPKVKKEVG-- 917
DB 1671 IKEYILVKABPKKERLLADQIESYSQOTD--QSVGIKTYLNNEIDPINPERSIRVATP 1728
QY 918 ---ADVYQLKEH-----YAGPDDNACKNQEVIRSSRTI-YLKTSGSKKLEAD- 963
DB 1729 RLIISSVSNFNFYIIIIILITLVAISIIIFVHY-----IKKNSLVIGILISOGYPLQISL 1784
QY 964 --TIDQLNQAVQAPLG-----LQSFYLD-----TERFGVFQKLA----- 996
DB 1785 SMTVFAMFTAVFGVFGVYVGNLQGLMNVFSNFWTLERHTEKFSIVSLMAVAVFLPLG 1844
QY 997 -----TSLAVQHKKQEKTLPKKLNNDGYTLIHDKLUKPKVPIPOISSPEKDFEGKLNQ 1049
DB 1845 MSILIIILTSMLIRKKPIDLM-----SGITAI-----KVGS-----FSRINR 1882
QY 1050 NGQSONVNSTFGSIIIESPYSTNFQEDADLDQDQDDSRQGNNSLNDQEAAGLLKQKLAI 1109
DB 1883 LFSKRNVTKFSASLLVNSFEK-----LISLAFAI 1912
QY 1110 LLGNQFIQYYQONDKEIEFEIINVERKVSSELSFRVEFKLAKLTLEDNGKTIIRVLSDETWSL- 1168
DB 1913 TITSTIIMFGISINKVFSKSIDSTYKNRHSFKL--NLATPTLEGGPLTKLNYDEIKNSL 1970
QY 1169 -----IVNTTIETKTPMSAVEPFDTKWVEQYDPTPLAAKTKFVLKFKDQIPV 1217

Db 1971 YVPLGTPGSAIVDGYTFKPGWSNAING--DRSQKNGDPKDYL---PHVFAFSMWLIV 2025
Qy 1218 DGS--GNISDKWLASIPVHIQOMRLSLPVVVTIRELGLKTEQOQOQOQOQOQOQOQPKKA 1275
Db 2026 EGAIREFNLWLIYNSLP--DTQKFR--VDKIVSEMGLLBOTQPDSTSPY-----KL 2073
Qy 1276 VRKEBELEY-----NPKDEFILNPLTKAHL-----TSLNVLNNDPNYKIEDLVKIK 1324
Db 2074 IQNPENGNYGVSVDPKNYFYFDLQGVGRGFFYMRMNAIENTYNHPIITTSRDI-FAD 2132
Qy 1325 NEAGD--HQLASLRANNIKRLMNTPIITFADYNPF-----FYNEDWRSIDKLYNNKG 1375
Db 2133 DYGDPPNQNISLR--NV-----YRQFLVRAVYTKLHSENR-IAKLEERG 2177
Qy 1376 NVSSHQQOAGGNGQSLGLQRLNKNIKPETTPA-----LIAKL 1414
Db 2178 TQSPFPREDYLSGGVVPDREF-----ETYSVASANFEGTTRSLNIYGYKSDSKFIKI 2232
Qy 1415 DRNNTWL-----SNYSKIMIMPKY--LVERSIGVPMWSTGLDGYIGSEOTKDGTSSSSQ- 1467
Db 2233 DENDVLMQEIATFETKTIDGKKVYPVWVNTVSSKKYKLGIGSTFGVSLNKATRYSHQI 2292
Qy 1468 -QKGPDDFI-QALGLKNTYHCKLGLSIRIFDPGNELAKIKDASNKGEEKLLKSYDLF 1525
Db 2293 KOENFDNNVIFVVGINPTFINEFITSOEIVNEITDLNKI-----ALRENOEIF 2342
Qy 1526 KNYLNEYKSPKIAKGTWNIHPDQKEYPNQKLPENLVNLVLPQWVTVLYNSDFIT 1585
Db 2343 NGILTSRDKPAQLI--GSSSIYASGWP-----AITNPEFK--GNESAIYD 2385
Qy 1586 NLFVPEGSDRGSGTKL-KQVIOKQNNN 1613
Db 2386 AIPDESGLRVSGLNLSLSEQILRWLNSN 2414

RESULT 39
T28317
ORF MSV156 hypothetical protein - Melanoplus sanguinipes entomopoxvirus
C:Species: Melanoplus sanguinipes entomopoxvirus
C:Date: 21-Jan-2000 #sequence_revision 21-Jan-2000 #text_change 09-Jul-2004
C:Accession: T28317
R:Afonso, C.L.; Tushman, E.R.; Lu, Z.; Oma, E.; Kutish, G.F.; Rock, D.L.
J. Virol. 73, 533-552, 1999
A:Title: The genome of Melanoplus sanguinipes entomopoxvirus.
A:Reference number: Z20484; MUID:99102612; PMID:9847359
A:Accession: T28317
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1127 <AFO>
A:Cross-references: UNIPROT:Q9YVT6; EMBL:AF063866; NID:g4049647; PIDN:AAC97677.1; PID:94
C:Genetics:
A:Note: MSV156

Query Match 2.88; Score 270; DB 2; Length 1127;
Best Local Similarity 19.88; Pred. No. 0.00069;
Matches 250; Conservative 187; Mismatches 426; Indels 398; Gaps 62;

Qy 533 KEENSKA-----INNQELEDNDITERLPEN-----SPIQYOQENAGLQASPD 576
Db 22 KYENKRVSLDIINSLEYILNN---IKFSDKITNEIKYNNKIKVEKIFYMENQ----- 70
Qy 577 KPYMKDVQORVYLAK--SQIOELIKADYTKLAKLLSNRHTVNTISLRLEKOLFVNPR 634
Db 71 --PKINDYNIITQLYEVNNEINKIKEN-----KPPCKNPLNYITKKKLYIYDLYE 122
Qy 635 IPSRDEI---KAKFVLDKTE--KNKQWQIYSSASPVFQNKWSIFGYRYRLGLG---DPK 686
Db 123 EKKOKELVINTEQKQAVDKINDIKNNVNNHSDNETIITGTETIDILNKLLKLVSSDEK 182
Qy 687 QTIHELKVLQKAGLQEGEYENLSPNLEDLKNIRIKTLPFSQKNFSLSLDFNNYYD 746
Db 183 QLIEQIYKNINNKIEIFNIDNVQKEIN---KKQDELNKLDESKEF----- 227

Qy 747 GEIKAPFGLPLFLPKELRRSNSGSGSONSPWEOEIIISQFKDONLS-----NQDOLA 801
Db 228 --IKQOE-----ELNKTIIDK-----QEELIKKLANDKEINFNIDEKQKLLD 266
Qy 802 QFSTKI--WEKIIIGDENEFQNNRLQYKLLKDJQESWINKTRNLNLYWYTGDKUKV--- 855
Db 267 QINSKINTLNENINGVMNLYTETK---NKISNLQNEILNK--DSTIKS-LDEKQKLLDE 319
Qy 856 --KPKNNLEAFR---QISNLOELLTAFYTSAAASNWNWYQDSGAKSTIIFEBIAELD 909
Db 320 LDKNNINNTSYLNKSNKTITNIQQLLESLLD---FNNANI-----NINELK 363
Qy 910 PKVKEKGVADVQLKFHYAIGFDNAGKFNQEVIRSSSR--TIYLTGSG---KSKLEAD--- 963
Db 364 SKIKL-----FDNDIQKLANDITEQNKKITDPFNNSTRIFKEKLDTEYK 407
Qy 964 TIDQLNOAVQNAPLQSLQSYLDTERFPGVQKATSLAVOHKQKE--KTLPKKLANDGYTL 1021
Db 408 KIDD-----IKNNNL-----QKLEESYKKIDEQTEYKKNKINKEYNDIIE 448
Qy 1022 IHDKLUKPVIPQISSSPKDNFEGKLNQNGSQNVNVTFGSIIIESPYFTNFQEDADLD 1081
Db 449 KNNNLQKLEENKKIDEQTEYKKNKINKE-----YNDIIEUK--NNNLQKLEEN 496
Qy 1082 ODGQDSDSGGNSLNDQEGALLKQKLAJLLGNQFOITYOQNDKEI-----EPE----- 1130
Db 497 KVINDKLTKLNDIESNTE-----LFLKNLISDFKDKSREIAKLNTEYEQLRKDL 547
Qy 1131 INVEKVSLSFRVEFKLA--KTLNENGTIRVLSDETMSLIVNTTIEKTPMSAVPEVD 1188
Db 548 ENINTNMLKMLDNKLSLEQLYDSKKNIL---DGIDKIYNSLKEKNDK---IDEYFS 600
Qy 1189 TKWVOYDPTPLAAKTKFVLKFKDQIPVDGSGNISDKWLASIPLVHQQMLRLSPVVK 1248
Db 601 N--IEKFDIYNVI--ENKFI-----GNLD-----SLINKI 626
Qy 1249 IRELGLKTEQOQOQOQOQOQOQKAVRKKEELETYNPKDEFNLTPLTKAHLRLSLNL 1308
Db 627 I-----NNDQFKEYINSKIDSKNELSTMF-DDIFNAKQOIA-----SITNN 667
Qy 1309 VNNDPNYKIEDLK--VIRKNEAGDHQALFSLRA-----NNIKRLMNTPI--TFAD----- 1353
Db 668 IENISN-KIKOLNEFIISNEDSSKELLDEIRKYKQOFDKIKDAMNTEVKSFENTLOKID 726
Qy 1354 -----YNPFYFNEDWRSIDKYLNKNKQNVSSHQOQOAGNCGSLIQRLNKNIKPE 1404
Db 727 SIKSNINELTNAYDIINTKANLDDKLNLYDSSEPKNLY-----NNASDLDDTIQKN--- 777
Qy 1405 TFTPALLALKDRNNTNLSNYSOKIIMIKPKVLVERSIGVPMWSTGLDGYIGSEOTKDGTS 1464
Db 778 -----NDEKVKOLNLEYLKE-----NKNQ 795
Qy 1465 SSQOKGFDDQFIQAL--GLANTEYHKGKLGSLIRIFDPGNELAKIKDASNKGEEKLLKSYD 1523
Db 796 SIEINDIVNNFIKELIKFNNTETNKSINELLNTDNDINDIKFELYKELNKISTNNLLK--- 852
Qy 1524 LFNQVLYNEYKSPKIAKGTWNIHPDQKEYPNP-----NOKLPENYLNVLNLPQWVTL 1577
Db 853 IYKNEIDNVNEK-----LSIVIENTLOFINSFLSIEFNQGSITSINFLN-----TL 899
Qy 1578 YNNSDFIT--NLFVPEGSDRGSGTKLQVIOKQNNVNNYADWGSAYLTFWYDKNIITNP 1635
Db 900 AGINDVNLKLNLIKIMADTTFRGD--TNIRDEIKNQIISSE----- 936
Qy 1636 NVITANIADVFIKDVKEL--EDNTKLIAPNITOWMPN-----ISGSKEKPY 1679
Db 937 NIKSQKFNEKNEKDLKLLISFNDKLNKYNISAGYTEYNNIEHCLKYLVIASDQBYRY 996
Qy 1680 K 1680
Db 997 K 997

[illegible]

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 28, 2005, 21:06:45 ; Search time 224 Seconds
(without alignments)
3506.709 Million cell updates/sec

Title: US-10-607-631-8
Perfect score: 9732
Sequence: 1 MNKKSTLLATATAAIIIGST.....TNNAFNNVFKFNISKIVE 1879

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1865214 seqs, 418043040 residues

Total number of hits satisfying chosen parameters: 1865214

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications AA:*

- 1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
- 2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep.*
- 3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep.*
- 4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep.*
- 5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep.*
- 6: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep.*
- 7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep.*
- 8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep.*
- 9: /cgn2_6/ptodata/1/pubpaa/US09A_PUBCOMB.pep.*
- 10: /cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pep.*
- 11: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep.*
- 12: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep.*
- 13: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep.*
- 14: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep.*
- 15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep.*
- 16: /cgn2_6/ptodata/1/pubpaa/US10D_PUBCOMB.pep.*
- 17: /cgn2_6/ptodata/1/pubpaa/US10E_PUBCOMB.pep.*
- 18: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep.*
- 19: /cgn2_6/ptodata/1/pubpaa/US11A_PUBCOMB.pep.*
- 20: /cgn2_6/ptodata/1/pubpaa/US11_NEW_PUB.pep.*
- 21: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep.*
- 22: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	9732	100.0	1879	15	US-10-607-631-8
2	9435.5	97.0	1878	15	US-10-607-631-20
3	475	4.9	1032	15	US-10-607-631-4
4	415	4.3	1001	15	US-10-607-631-10
5	370.5	3.8	1939	17	US-10-732-923-3340
6	350	3.6	2910	17	US-10-732-923-3342
7	346	3.6	1985	17	US-10-732-923-3351
8	325	3.3	4688	15	US-10-282-122A-76865
9	305.5	3.1	1875	15	US-10-369-493-22285
10	305.5	3.1	1875	17	US-10-732-923-3334
11	305.5	3.1	1875	17	US-10-732-923-3335

12	305.5	3.1	2503	17	US-10-828-985A-11	Sequence 11, Appl
13	305.5	3.1	2543	17	US-10-828-985A-9	Sequence 9, Appl
14	305.5	3.1	2568	17	US-10-828-985A-7	Sequence 7, Appl
15	303.5	3.1	1441	17	US-10-732-923-3352	Sequence 3352, Ap
16	303	3.1	2633	18	US-10-450-763-36864	Sequence 36864, A
17	301	3.1	2663	16	US-10-723-860-749	Sequence 749, App
18	300	3.1	1527	17	US-10-732-923-3354	Sequence 3354, Ap
19	299	3.1	2017	18	US-10-450-763-36660	Sequence 36660, A
20	297	3.1	1404	17	US-10-732-923-3304	Sequence 3304, Ap
21	295	3.0	1602	17	US-10-954-924-7	Sequence 7, Appl
22	295	3.0	1621	17	US-10-954-924-5	Sequence 5, Appl
23	295	3.0	1639	14	US-10-087-464-10	Sequence 10, Appl
24	295	3.0	1639	17	US-10-954-924-3	Sequence 3, Appl
25	291	3.0	5005	15	US-10-282-122A-76871	Sequence 76871, A
26	290.5	3.0	1679	15	US-10-369-493-22080	Sequence 22080, A
27	288	3.0	1948	14	US-10-032-585-7611	Sequence 7611, Ap
28	286.5	2.9	1979	14	US-10-205-823-419	Sequence 419, App
29	286.5	2.9	1979	20	US-11-051-454-419	Sequence 8087, Ap
30	283.5	2.9	2384	15	US-10-335-977-8087	Sequence 73, Appl
31	281	2.9	1805	10	US-09-820-843A-73	Sequence 63513, A
32	281	2.9	1805	15	US-10-282-122A-63513	Sequence 3303, Ap
33	281	2.9	1805	17	US-10-732-923-3303	Sequence 64245, A
34	280.5	2.9	1818	15	US-10-282-122A-64245	Sequence 3306, Ap
35	280.5	2.9	1818	17	US-10-732-923-3306	Sequence 8088, Ap
36	280.5	2.9	2440	15	US-10-335-977-8088	Sequence 3353, Ap
37	279.5	2.9	1478	17	US-10-732-923-3353	Sequence 23, Appl
38	275.5	2.8	10203	16	US-10-661-809-23	Sequence 4098, Ap
39	275.5	2.8	10203	18	US-10-724-972A-4098	Sequence 8312, Ap
40	274.5	2.8	1997	17	US-10-732-923-8312	Sequence 10224, A
41	273.5	2.8	1690	20	US-11-097-143-10224	Sequence 10311, A
42	273.5	2.8	1690	20	US-11-097-143-10311	Sequence 254, App
43	271	2.8	3225	16	US-10-408-765A-254	Sequence 70580, A
44	270	2.8	6641	15	US-10-282-122A-70580	Sequence 17015, A
45	267	2.7	2237	17	US-10-732-923-17015	

ALIGNMENTS

RESULT 1
US-10-607-631-8
; Sequence 8, Application US/10607631
; Publication No. US20040091901A1
; GENERAL INFORMATION:
; APPLICANT: Minion, F. Chris
; APPLICANT: Djordjevic, Steven P.
; TITLE OF INVENTION: Immunogenic Mycoplasma Hyopneumoniae
; TITLE OF INVENTION: Polypeptides
; FILE REFERENCE: 08411/035001
; CURRENT APPLICATION NUMBER: US/10/607,631
; CURRENT FILING DATE: 2003-06-27
; PRIOR APPLICATION NUMBER: 60/392,632
; PRIOR FILING DATE: 2002-06-28
; NUMBER OF SEQ ID NOS: 42
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 8
; LENGTH: 1879
; TYPE: PRT
; ORGANISM: Mycoplasma hyopneumoniae
US-10-607-631-8

Query Match 100.0% Score 9732; DB 15; Length 1879;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1879; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy	1	MNKKSTLLATATAAIIIGSTVFGTVVGLASKVKYGVNPTQGVISQLGLIDSVAFKPSIA	60
Db	1	MNKKSTLLATATAAIIIGSTVFGTVVGLASKVKYGVNPTQGVISQLGLIDSVAFKPSIA	60
Qy	61	NFTSDYQSVKALLNGKTFDPKSSBFTFVSKFDELTNNGRTVLEIPKKYQVWISFESPE	120
Db	61	NFTSDYQSVKALLNGKTFDPKSSBFTFVSKFDELTNNGRTVLEIPKKYQVWISFESPE	120

QY 121 DDKERFRLGPHLKEKJEDGNIAQSATKFIYLLPLDMPKAAALGOYSYIVDKQFNLLIHL 180
DB 121 DDKERFRLGPHLKEKJEDGNIAQSATKFIYLLPLDMPKAAALGOYSYIVDKQFNLLIHL 180
QY 181 SNFSAOSIKPLATRRSDFTAKLNQFNQDELAVYLEKFPDLALKANIRLOTADESFEK 240
DB 181 SNFSAOSIKPLATRRSDFTAKLNQFNQDELAVYLEKFPDLALKANIRLOTADESFEK 240
QY 241 GNLVDPPVYSFIRNPQKQEWASDLNQDKTVRLYLRTFSPQAKTILKDYKYKDTFTLS 300
DB 241 GNLVDPPVYSFIRNPQKQEWASDLNQDKTVRLYLRTFSPQAKTILKDYKYKDTFTLS 300
QY 301 SIDLKASNGTSLPANENDLKQDLVDLLVSDVFGQSETITSNSQVQKVPASERSLKDR 360
DB 301 SIDLKASNGTSLPANENDLKQDLVDLLVSDVFGQSETITSNSQVQKVPASERSLKDR 360
QY 361 VKPKDQOKPRIEKFSLEYDALSFYSOLOELYSKNSIKDLVNATLARNLRESLCKYNF 420
DB 361 VKPKDQOKPRIEKFSLEYDALSFYSOLOELYSKNSIKDLVNATLARNLRESLCKYNF 420
QY 421 LFDLASHLDYYFLVSKAKIKQSSITKLPFIEPIKISLSKSSILGQEPNIKTLFEKEVT 480
DB 421 LFDLASHLDYYFLVSKAKIKQSSITKLPFIEPIKISLSKSSILGQEPNIKTLFEKEVT 480
QY 481 FKLDNFRDVEIEKAFGLLYPGVNEELEQAKAQRASFEKEKSKGKJEFSSQKKEENSKAI 540
DB 481 FKLDNFRDVEIEKAFGLLYPGVNEELEQAKAQRASFEKEKSKGKJEFSSQKKEENSKAI 540
QY 541 NNQEGLEEDNITERLPENSPIOYOQENAGLGASDPKPYMIKDVQNRYYLAKSQIQELI 600
DB 541 NNQEGLEEDNITERLPENSPIOYOQENAGLGASDPKPYMIKDVQNRYYLAKSQIQELI 600
QY 601 KADYTKLAKLLSNRHTYINSLRKEQLFDVNPRIIPSSRDIEKAKFVLDKTERNKYQWY 660
DB 601 KADYTKLAKLLSNRHTYINSLRKEQLFDVNPRIIPSSRDIEKAKFVLDKTERNKYQWY 660
QY 661 SSASPVPQKWSLFGYYRYLLGLDPKQTIHELKVLGQKAGLQPEGYENILPSDPNLEDLKN 720
DB 661 SSASPVPQKWSLFGYYRYLLGLDPKQTIHELKVLGQKAGLQPEGYENILPSDPNLEDLKN 720
QY 721 IRIKTPFSQKDNFKLSLLDFNNYDGEIKAPFGLPLPLKELRNSNSGGSQNSNP 780
DB 721 IRIKTPFSQKDNFKLSLLDFNNYDGEIKAPFGLPLPLKELRNSNSGGSQNSNP 780
QY 781 WEQEIIISQFQDNLSNQDLAQFSTKIWEKIIIDENEFDQNNRLOKLLKDLQESWINKT 840
DB 781 WEQEIIISQFQDNLSNQDLAQFSTKIWEKIIIDENEFDQNNRLOKLLKDLQESWINKT 840
QY 841 RDNLWYTYLGDKLKVPKNNLEAKFRQISNLQELLTAFYTSAAALSNWNYQDSGAKSTI 900
DB 841 RDNLWYTYLGDKLKVPKNNLEAKFRQISNLQELLTAFYTSAAALSNWNYQDSGAKSTI 900
QY 901 IFEIEIAELDPKVEKVGADYVQLKFHYAIGFDDNAGKFNQEVIRSSSRITLYLKTSKSKL 960
DB 901 IFEIEIAELDPKVEKVGADYVQLKFHYAIGFDDNAGKFNQEVIRSSSRITLYLKTSKSKL 960
QY 961 EADTIDQLNQAVKNAPLGLSQFYLDTFRFGVFOKLATSLAVQHKQEKTLPKKLNDGYT 1020
DB 961 EADTIDQLNQAVKNAPLGLSQFYLDTFRFGVFOKLATSLAVQHKQEKTLPKKLNDGYT 1020
QY 1021 LIHDKLKVPVIPOISSPEKOWFEGKLNQNGQSONVSTFGSIIIESPYFSTNFQDADL 1080
DB 1021 LIHDKLKVPVIPOISSPEKOWFEGKLNQNGQSONVSTFGSIIIESPYFSTNFQDADL 1080
QY 1081 DQGGQDSSROGNSLNDQEBAGLLKQKLAILLGNQFIQYQOQNDKEFEFELINVEKYSLS 1140
DB 1081 DQGGQDSSROGNSLNDQEBAGLLKQKLAILLGNQFIQYQOQNDKEFEFELINVEKYSLS 1140
QY 1141 FRVEFKLAKTLEDNGKTIIRVLSDETMSLI VNTTIEKTPEMSAYPEVFDTKWBEQYDPRTP 1200
DB 1141 FRVEFKLAKTLEDNGKTIIRVLSDETMSLI VNTTIEKTPEMSAYPEVFDTKWBEQYDPRTP 1200
QY 1201 LAAKTFVLKFKDQIPVDGSGNISDKWLASIPLVIHQMLRLSPVVKTTIRELGLKTEQOQ 1260

DB 1201 LAAKTFVLKFKDQIPVDGSGNISDKWLASIPLVIHQMLRLSPVVKTTIRELGLKTEQOQ 1260
QY 1261 QOQOQOQOQOQOQKAVRKEELETYNPKDEFNLANPLTKAHLTLTSLNVNNDPNYKIEDL 1320
DB 1261 QOQOQOQOQOQOQKAVRKEELETYNPKDEFNLANPLTKAHLTLTSLNVNNDPNYKIEDL 1320
QY 1321 KVIKNEAGDHQALFASLRANNIKRLMNTPIITADYNPFFYYNEDWRSIDKYLANKGNVSSH 1380
DB 1321 KVIKNEAGDHQALFASLRANNIKRLMNTPIITADYNPFFYYNEDWRSIDKYLANKGNVSSH 1380
QY 1381 QOQAAAGGOGSGLIQLRLNKNIKPETFPTPALIALKORNTNLSNYSDKIIMIKPYLVERS 1440
DB 1381 QOQAAAGGOGSGLIQLRLNKNIKPETFPTPALIALKORNTNLSNYSDKIIMIKPYLVERS 1440
QY 1441 IGVPMSTGLDGYIGSEQTKDGTSSSQKGFQDDFTIQLGLKNTYHGLGSLIRIFDPG 1500
DB 1441 IGVPMSTGLDGYIGSEQTKDGTSSSQKGFQDDFTIQLGLKNTYHGLGSLIRIFDPG 1500
QY 1501 NELAKIKDASNKKGBEKLKSYDLFKNYLNHYEKKSPKIAKGWTNIHPDQKYPNPQKL 1560
DB 1501 NELAKIKDASNKKGBEKLKSYDLFKNYLNHYEKKSPKIAKGWTNIHPDQKYPNPQKL 1560
QY 1561 PENYLNVLNQPKWKTLYNSSDFTNLVFEPSGRSGCTKLKQVLOKOVNNYADWMSA 1620
DB 1561 PENYLNVLNQPKWKTLYNSSDFTNLVFEPSGRSGCTKLKQVLOKOVNNYADWMSA 1620
QY 1621 YLTFWYDKNIIINQPNVITANIADYFIKDVLEEDNTKLIAPNITQMWPNISGSKEKFKY 1680
DB 1621 YLTFWYDKNIIINQPNVITANIADYFIKDVLEEDNTKLIAPNITQMWPNISGSKEKFKY 1680
QY 1681 PTVFFGNWENENSNMSQAQTTWEKIREGFALQALKSSFDQKTRTFVLTNNAPLPLWKY 1740
DB 1681 PTVFFGNWENENSNMSQAQTTWEKIREGFALQALKSSFDQKTRTFVLTNNAPLPLWKY 1740
QY 1741 GPLGQNGENFTQDMRLVFQNDNQIALRALRVOEDRBEKSEDDKQKWKFKVPIPEE 1800
DB 1741 GPLGQNGENFTQDMRLVFQNDNQIALRALRVOEDRBEKSEDDKQKWKFKVPIPEE 1800
QY 1801 MFNSGNIRFVGVMQIQGPNLWLPVINSVIYDFYRGTDSDNDVANLNVAPQVKTIAPT 1860
DB 1801 MFNSGNIRFVGVMQIQGPNLWLPVINSVIYDFYRGTDSDNDVANLNVAPQVKTIAPT 1860
QY 1861 NNAFNNVKEFNISKIIVE 1879
DB 1861 NNAFNNVKEFNISKIIVE 1879
RESULT 2
US-10-607-631-20
; Sequence 20, Application US/10607631
; Publication No. US20040091901A1
; GENERAL INFORMATION:
; APPLICANT: Minion, F. Chris
; APPLICANT: Djordjevic, Steven P.
; TITLE OF INVENTION: Immunogenic Mycoplasma Hyopneumoniae
; FILE OF INVENTION: Polypeptides
; FILE REFERENCE: 08411/035001
; CURRENT APPLICATION NUMBER: US/10/607,631
; CURRENT FILING DATE: 2003-06-27
; PRIOR APPLICATION NUMBER: 60/392,632
; PRIOR FILING DATE: 2002-06-28
; NUMBER OF SEQ ID NOS: 42
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 20
; LENGTH: 1878
; TYPE: PR1
; ORGANISM: Mycoplasma hyopneumoniae
US-10-607-631-20
Query Match 97.0%; Score 9435.5; DB 15; Length 1878;
Best Local Similarity 97.0%; Pred. No. 0;
Matches 1827; Conservative 17; Mismatches 29; Indels 11; Gaps 3;

QY 1 MNKKSTLLATAAAIIGSTVGTWGLAGKVKYGVNPTQGVISQGLIDSVAFKPSIA 60
DB 1 MNKKSTLLATAAAIIGSTVGTWGLAGKVKYGVNPTQGVISQGLIDSVAFKPSIA 60
QY 61 NPTSDYQSVKALLNGKTFDPKSSEFTDFVSKEDPLTNGRTVLIEIPKKYQVVISSEPS 120
DB 61 NPTSDYQSVKALLNGKTFDPKSSEFTDFVSKEDPLTNGRTVLIEIPKKYQVVISSEPS 120
QY 121 DDKERFRLGLHKEKLEJEDGNIQAQATKFIYLLPLDMPKAALGOYSYIVDKNFNNLIHPL 180
DB 121 DDKERFRLGLHKEKLEJEDGNIQAQATKFIYLLPLDMPKAALGOYSYIVDKNFNNLIHPL 180
QY 181 SNFSAQSIKPLALTRSSDFIAKLNOFNQDELWYLEKFPDLEALKANIRLOTADPSFEK 240
DB 181 SNFSAQSIKPLALTRSSDFIAKLNOFNQDELWYLEKFPDLEALKANIRLOTADPSFEK 240
QY 241 GNLVDPFVYSFIRPQOKEWASDLNODQKTVRLYLRTFSPQAKTILKDYKYKDETFLS 300
DB 241 GNLVDPFVYSFIRPQOKEWASDLNODQKTVRLYLRTFSPQAKTILKDYKYKDETFLS 300
QY 301 SIDLKASNGTSLFANENDLKQDLDVLDVSDYFGGQSETITTSNQVPPASERSLKDR 360
DB 301 SIDLKASNGTSLFANENDLKQDLDVLDVSDYFGGQSETITTSNQVPPASERSLKDR 360
QY 361 VKFKKQOKPRIKPSLEYDALSFYSQLOBLVSKPNSIKOLVNATLARNLRFSLGKYNF 420
DB 361 VKFKKQOKPRIKPSLEYDALSFYSQLOBLVSKPNSIKOLVNATLARNLRFSLGKYNF 420
QY 421 LPDDLASHLDYFLVSKAKIKQSSITTKLFIELPIKISLKSIIIGDQEPNITLPEKEVT 480
DB 421 LPDDLASHLDYFLVSKAKIKQSSITTKLFIELPIKISLKSIIIGDQEPNITLPEKEVT 480
QY 481 FKLDNFRDVEIEKAFGLLYPGVNELEQARQAQASFEKSKSCKGLKEFSQCKEENSKAI 540
DB 481 FKLDNFRDVEIEKAFGLLYPGVNELEQARQAQASFEKSKSCKGLKEFSQCKEENSKAI 540
QY 541 NNQGLEBDDNITERLPENSPYQOENAGLQSPDKPYMIKDVQNYRYLAKSQIQBELI 600
DB 541 NNQGLEBDDNITERLPENSPYQOENAGLQSPDKPYMIKDVQNYRYLAKSQIQBELI 600
QY 601 KAKDYTKLAKLSNRTYNIISLKEOLFVNPRIPSSRDIEKAFVLDKTEKNKYWIY 660
DB 601 KAKDYTKLAKLSNRTYNIISLKEOLFVNPRIPSSRDIEKAFVLDKTEKNKYWIY 660
QY 661 SSASPVFQNKSLFGYRYLLGLDPKQTIHELVLKQKAGLQFEGYENLPSDFNLEDLN 720
DB 661 SSASPVFQNKSLFGYRYLLGLDPKQTIHELVLKQKAGLQFEGYENLPSDFNLEDLN 720
QY 721 IRIKTPLFQKDNFKLSLLDFNNYDGEIKAPEFGLPLPKELRRNSSGGSSQNSNP 780
DB 721 IRIKTPLFQKDNFKLSLLDFNNYDGEIKAPEFGLPLPKELRRNSSGGSSQNSNP 780
QY 781 WEQEIISQKQNLNODLOAQFSTKIWEKIIIGDSEFPDQNNRLOVLLKDLQESWINKT 840
DB 781 WEQEIISQKQNLNODLOAQFSTKIWEKIIIGDSEFPDQNNRLOVLLKDLQESWINKT 840
QY 841 RNLNWTYLLGDLKVKPKNNLEAKFRQISNLOELLTAFYTSAAALSNNMNYQDSGAKSTI 900
DB 841 RNLNWTYLLGDLKVKPKNNLEAKFRQISNLOELLTAFYTSAAALSNNMNYQDSGAKSTI 900
QY 901 IFEETIABLPKVKEGADVYQLKHYAIGFDNAGKFNQEVIRSSRTIYLKTSQKSKL 960
DB 901 IFEETIABLPKVKEGADVYQLKHYAIGFDNAGKFNQEVIRSSRTIYLKTSQKSKL 960
QY 961 EADTDOLNQAQNPAGLQSFVLDTEREGVFKLQSLATSLAVOHKQKTEPLPKLNDGVT 1020
DB 961 EADTDOLNQAQNPAGLQSFVLDTEREGVFKLQSLATSLAVOHKQKTEPLPKLNDGVT 1020
QY 1021 LIHDKLKXVPIQIISSEPEKOWPEGLKNGQSQNVNVTFGSIIESPYFSTNFQEDADL 1080
DB 1021 LIHDKLKXVPIQIISSEPEKOWPEGLKNGQSQNVNVTFGSIIESPYFSTNFQEDADL 1080

QY 1081 DQDQDDSRQGNNSLDNQEAGLLKQKLAJLLGNQFIQYQOQDKIEFEIINVEKYSLS 1140
DB 1081 DQDQDDSRQGNNSLDNQEAGLLKQKLAJLLGNQFIQYQOQDKIEFEIINVEKYSLS 1140
QY 1141 FRVEFKLAKTLEDNGKTIIRVLSDETMSLIWNTTIEKTPMSAVPEVDFTKWQYDPRTP 1200
DB 1141 FRVEFKLAKTLEDNGKTIIRVLSDETMSLIWNTTIEKAPMSAAPEVDFTKWQYDPRTP 1200
QY 1201 LAAKTFVLKFKDOIIPVDSGNISDKWLASIPLVIHQOMLRLSPVVKTIRELGLKTEQQQ 1260
DB 1201 LAAKTFVLKFKDOIIPVDSGNISDKWLASIPLVIHQOMLRLSPVVKTIRELGLKTEQQQ 1260
QY 1261 QOQOQOQOQOQKAVRKEBELETYNPKOBFNLIPLTKAHLRTLSLNVNNDPNYKIIDL 1320
DB 1261 QOQOQOQOQOQKAVRKEBELETYNPKOBFNLIPLTKAHLRTLSLNVNNDPNYKIIDL 1320
QY 1321 KVIKNEAGDHLAFSLRANNIKRLMNTPTTFADYNPFYFYNEDWRSIDKYLNNKGNVSS- 1379
DB 1321 KVIKNEAGDHLAFSLRANNIKRLMNTPTTFADYNPFYFYNEDWRSIDKYLNNKGNVSS- 1379
QY 1379 KVIKNEAGDHLAFSLRANNIKRLMNTPTTFADYNPFYFYNEDWRSIDKYLNNKGNVSS- 1379
DB 1379 KVIKNEAGDHLAFSLRANNIKRLMNTPTTFADYNPFYFYNEDWRSIDKYLNNKGNVSS- 1379
QY 1380 ----HQOQOAGGNGSGLIQRLAKNPKETFTPALIALKDRNNTNLSNYSDKIIMIKPY 1435
DB 1380 ----HQOQOAGGNGSGLIQRLAKNPKETFTPALIALKDRNNTNLSNYSDKIIMIKPY 1435
QY 1435 LVERSIGVPWSTGLDGYIGSEOTKQGTSSSSQKGFQDQFIQALGLKNTYHGLGLSIR 1495
DB 1435 LVERSIGVPWSTGLDGYIGSEOTKQGTSSSSQKGFQDQFIQALGLKNTYHGLGLSIR 1495
QY 1495 IFDPGNSLAKIKOASNNKGBEKLKSYDLPKNYLNEYEKSPKIAKGWTHIHDPQKEYPN 1555
DB 1495 IFDPGNSLAKIKOASNNKGBEKLKSYDLPKNYLNEYEKSPKIAKGWTHIHDPQKEYPN 1555
QY 1555 PNQKLPENYLNLVNOQPKVTLVNSSDPITNLVPEPGSDRGSGTKLQVIOKOVNNYA 1615
DB 1555 PNQKLPENYLNLVNOQPKVTLVNSSDPITNLVPEPGSDRGSGTKLQVIOKOVNNYA 1615
QY 1615 DWGSAYLTFWYDKNIIINQPNVITANITADVFIKDVKELEDNTKLIAPNITQWPNISGSK 1675
DB 1615 DWGSAYLTFWYDKNIIINQPNVITANITADVFIKDVKELEDNTKLIAPNITQWPNISGSK 1675
QY 1675 EKPYKPTVFFGNMENNENSNMSQOQTPTWBKIREGFPALQALKSSFDQKTRTFVLTNNAPL 1735
DB 1675 EKPYKPTVFFGNMENNENSNMSQOQTPTWBKIREGFPALQALKSSFDQKTRTFVLTNNAPL 1735
QY 1735 PLWKYGLPGFQNGPNFKTQDWRLVFQNDNQIARVQEOQDRPEKSSDKQKWKIPKV 1795
DB 1735 PLWKYGLPGFQNGPNFKTQDWRLVFQNDNQIARVQEOQDRPEKSSDKQKWKIPKV 1795
QY 1795 VIPEMFNSGNIRFVGWQIQGPNTLWLPVINSSVIYDFYRGTDGSDNDVANLNAVQVK 1855
DB 1795 VIPEMFNSGNIRFVGWQIQGPNTLWLPVINSSVIYDFYRGTDGSDNDVANLNAVQVK 1855
QY 1856 TIAFTNNAFNVKPEFNISKIIVE 1879
DB 1856 TIAFTNNAFNVKPEFNISKIIVE 1879

RESULT 3

US-10-607-631-4
; Sequence 4, Application US/10607631
; Publication No. US20040091901A1
; GENERAL INFORMATION:
; APPLICANT: Minion, F. Chris
; APPLICANT: Djordjevic, Steven P.
; TITLE OF INVENTION: Immuno-genetic Mycoplasma Hyopneumoniae
; TITLE OF INVENTION: Polypeptides
; FILE REFERENCE: 08411/035001
; CURRENT APPLICATION NUMBER: US/10/607,631
; PRIOR FILING DATE: 2003-06-27
; PRIOR APPLICATION NUMBER: 60/392,632
; NUMBER OF SEQ ID NOS: 42
; SOFTWARE: FastSeq for Windows Version 4.0


```
; SEQ ID NO 76865
; LENGTH: 4688
; TYPE: PRT
; ORGANISM: Ureaplasma urealyticum
US-10-282-122A-76865

Query Match      3.3%; Score 325; DB 15; Length 4688;
Best Local Similarity 19.4%; Pred. No. 1e-08;
Matches 450; Conservative 309; Mismatches 771; Indels 786; Gaps 118;

Qy 40  TQGVISOGLGLDSVAFKPS-IANFTSDYQSVKALLNGKTFDPKSSSEFTDVSFDFLTN 98
Db   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 785  TKTVEASIGLOGKAIFKTSDDAIAPDH---KYTLTKIEADNKVANIDEISPLDRIVN 840
Db   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 99   ---NGRTVLEIPKKYQVVISFSPED---DKERFLGHFLKKEKLED-GNTAQSAKTFIY 150
Db   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 841  KQKGN--VNADNKEHFKIPQKNKDLTAVYKDKNNIEHPIKTDGKGVIVPNPNLNF 898
Db   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 151  LPLDMPK-----AALGOY--SVIVDKNFN-----LIHPLSNFS 184
Db   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 899  ---DPNKIYDFDKIVDLNEYPNKTILDRNSINKDVSAINDGVDOQARKLVKAPAVSNVT 954
Db   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 165  AQSI-----KPLALTRSSDFTAKNOFNQDBELWYLEKF----- 219
Db   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 955  VNAINFQVNLFDLNLKLSYNRQFALT-----IAKVNDLNDTQK---YIATYDPKTNKYKLN 1006
Db   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 220  FDEALKANIRLQTADFSFEKGNLVDPPVYGFIRNPQKQWASDLNODQKTURL-----Y 275
Db   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 1007  FDFTHLDANTKYKVDVEL-----LELNKKEPKIKLIKDDV 1042
Db   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 276  LRTESPQAKTILK-----DYKYK---DETFLSSIDLKASGTSLFANENDLKQD-- 323
Db   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 1043  LNFETTSATTINPPIWTKFVWVTKNNDTTITFEIDDK-----DNILNKQKIYAQUAL 1098
Db   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 324  -DVDLLDVSDFRG---GQSETITSNQKVPASERSLDRVAFKKDQKQPRIEKSFL 377
Db   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 1099  MDDLLD--TDVINPLVNSFKTNKIASINGLDLKGNKYKSIKLYLNDQNKVYLFKNDV 1157
Db   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 378  YEYDALSFYQIQELVSKPNSIKDLVNATLARNLPSLGKYNFLFDDLAHLDDYFLVSK 437
Db   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 1158  TKYE-----QHEFTNPHKI-----NLSFN--KSAVEQDIFADHANLF----- 1192
Db   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 438  AKIKOSSITKFLFELPIKI-----SLKSSILGDQEBNIKTL 474
Db   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 1193  --IDYKDYDQKLRIINEDVKIYYQNIIDNTKNELOIGYGVKVANNTKFLNLVGLKE---KTT 1247
Db   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 475  FEKEVTFKLDNFDROVEIEKAFGLLYPGVNEELEQAKQARASFEKEKSKGLKEFSQOKE 534
Db   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 1248  YVKKLEALNKSASSIVNSEFDLLDPNTN-----FSTSNGKNTTIVGLNSIDNWN 1297
Db   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 535  ENSKAINNOGLEED--DNITERLPENSPIOYQOENAGLGASDPKPYMIKQVONORYILA 592
Db   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 1298  DQTPPIAKINIGDDFQDNOQVKL-----IYVSDNKEI-----KSKAVTLIKGQRY-- 1345
Db   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 593  KSQIQELIYAKDYTKLAKLL--SNRHTYNISRLKEQLFDVNP-----RIPSSRDIEKAK 645
Db   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 1346  QFEFSNLIKRLYT--FSKIVYETNNQTLKLDLTLHQ--FSINPSNNAVSLKNTNIEITK 1403
Db   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 646  FVL-----DKTEKNKYQIYSSASPVFQNKW 671
Db   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 1404  RILVNDQSLISAKIEVDIDIVLNTADKPNIVYQLENSDANNLKLATNAQVVENNK 1463
Db   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 672  SLFGYYRYLLGLDPKQTIHELKVLGQAGLQFEGYENLPSD--FNLLEDKNIRIKTPLFS 729
Db   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 1464  KFLKFDLSLKINQNVVKEI-----SFNSKPTNAYFNFTNNKTNNI---VYS 1508
Db   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 730  QKDNFKLSI-----LDFNNYDGEIKAPFGLPLPKELRN 767
Db   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 1509  YDEQNKISLSNNINPTSYKPTQNTKNDKSVNIDVDLQVDKLLANQY--LRLKL--KOLND 1566
Db   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 768  SS-----NSGGSON-----SNSPWEQEIISQFDQNLNQ--DQLAQFSTKIWEK 810
Db   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 1567  KTWVTDPILFNNNAKISPKLSNLIHNRAYELEGLYFFDQNSVNDMTNNQISFNSKIHKP 1626
Db   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

811  IIGDEN-----BFDONNRLOYKLLKDOBSWINKTRDNLWYTLGDKLVKPKNKL----- 861
1627  KIBFEPSLTTINTYDNTNNAIKTSAHNAQVHFUKLTNDEALENDQIIVBAVAPTNNLNDQK 1686
862  --BAKPRQISN-----LQELLTAFYTSAAALSNNNNYQDSGAKSTIIPBEIA 906
1687  VWEAKLNNVNSFNEGELEFNLGLREBETTVRLIKVTFKNKPKAKAYELLNKGVIPE--- 1743
907  ELDPKVKEKGVADYQKLYHAYIGFDDNAGKFNQEVIRSSRTIYLYKTSKSKLEATID 966
1744  -----YKNGSOAYEF-----TTQFHEKVIDVVSSTSTNTTQOEITVK---ID 1783
967  QLNQAVKNAPLGLH---QSFYLDTERFGVFOKLAATSLAVQHKQKELTKPKLNNNDGYTLIH 1023
1784  GIQARMNKKLELVESNIIIGDPEIKTTVDNNNSVHLSFDKKE-----YNLVL 1832
1024  DKLKPKVIPQITSSPEKDFEGKLN-----QNGSQS-----NVNVSTFGSIIESPYFSTN 1073
1833  NNLK-----PGRYSIKKINIKEDVNGQDHFVKEINVN--NSFDVNLQSEITASS 1881
1074  FOEDADLDQDQDSROGNSLNDQEAAGLLKQKLAILLNQ-----FIQYQOQNDKEIEF 1128
1882  VEEINDRAPQKLNQTTIKNLKDDNDILKTNDIATITYDNEQKVDIAIVKTNQAKQYLEA 1941
1129  EIIN-VEKVSLSFRVEFK-LAKTLEDNGK-TIRVLSDET--MSLIVN-----TT 1173
1942  IITLVFNKNVITIKIEFKNLSQTFIKVGKNTNVIYDESMLNKLIINNDPQIIGPLSTD 2001
1174  IEKTPEMSA-----VPBVFDTKWVEQYDPRTPLAAKTKFVLKFDQIPVDGSGNISDKWL 1228
2002  VNSTQNVANNKHVISSTLDFK-----VNPH--ISKNLKFKLKFQ-----NINGS-- 2044
1229  ASIPLVTHQMLRSLSPVVKTIRELGLKTEQOQOQOQOQOQOQOQOQOQOQOQOQOQO 1288
2045  -----VVSYPILTNSII-----VNNK 2062
1289  DEFNLAPLTKAHLTLISLVNNDPNYKIEDLKI-----KNEAGDHLAPSLRAN 1339
2063  ---NVIN-----FTLDNLSNQ-LRLVDVYVIDDNDTINDKXVP-----KAN 2103
1340  NIKELMN-----TPITFADYNPFYFNEBWRSIDKYLNNKGNVSSHQOQAGGQSGLI 1394
2104  NVTRIIDIAPKTTISKS-----NNTW-----NTTSSQPEFVINSDDGNEVL 2147
1395  QRLNKNI---KPET-FTPALIALKDRN-----TNLSNYS-----DKIIMIKP---K 1434
2148  DNLEATTISFKKGOTLLTPVKVNIKKONNKYLIKQOITNLEPENRYVLESILLAKPNKTK 2207
1435  YLVERSIG---VPWSTGLDGY---IGSE-----QTK-----DGTSSSQOQGFDDQF 1475
2208  PLVVEILNKDISFQTOAGNYKVQIKSQNPSTVDTQRIKLDGIGNANNEKQLEITY 2267
1476  -----IQALGLKNTEYHGKLG-----LSIRIFD-----PGNELAKIKD- 1508
2268  SANDNSTKTAIIKLEKNKLEYEFELTNLEKRTVTFTFKIELINDNNTKTTPFKSDSIQDK 2327
1509  ---ASNKKGBEKLLKSYDLFPNYLNE-----YKSKSPKIAK 1541
2328  FIVLSNNQVGVGNIIEIQDRDVRNHLNSAKIRFELNDLVNSDQEAATITVNNNTTSAK 2387
1542  GWTNIHPDQEPNPNQK-LPENYLVNLVNPWKVTLVNSSDFIINLFPVEBEGSDRGSGT 1600
2388  ---VITDQ-----NQKLEATFSNLVNL---KDTIINKIEFNT-----KPKNASKNIGI 2430
1601  KLKQVIOQRQ-----VNNNYADMGSAVLTFWYDKNIITN-----QPNVITANIADVFI- 1647
2431  NDTNVIYDATNLIINNDLKITGPLHTLKEFBANNNKTNISVLSLELDTNNHKNL---YFIA 2488
1648  -----KDVKELEDNTKLTAPNITQWPNISGSKEKFPKTFVFPFNWE 1689
2489  KFDSDNGOSVLTNPITAADKIVTNNNNNNKKELTFLNT-----NLTSNRQYIFGLXYVNSN 2544
```


Db 1645 DEGRQOAMMTTLERKLAKMESQLETKOABSPKSVNNVQPLGLPRKIBEN--SN 1702
Qy 1312 DP-NYKIEDLVKINEAGDHQAFSLRANNIKRLMNTPIITPADYNPFYFNEDWRSIDKY 1370
Db 1703 SPFNPLSGEKLKLNKSSSGGFN-----PFTSPSPKHLQNDNDKR---ES 1747
Qy 1371 LNNKGNVSHQOQAGNGQSGLI---ORLNKNIKPTFTPALIALKDRNTNLSYSDK 1427
Db 1748 LANKTDPPTLPEFPNIPASRGLISSSTLSTNDDELTSSNPAQKSSNRNVQSEED- 1806
Qy 1428 IIMIKPKLVERSICVPHSTGLDVGIGSEQTKDGTSSSQOKGFDQFIQALGLKNTYEH 1487
Db 1807 -----TEKKGEPE-----VKGAEIEEQTKSKRPIDEVGELKNDK-- 1843
Qy 1488 GKLGLSIRIPGNEKAKIDASKKGBEKLKSYDLFPNLYNEFKSPKI 1539
Db 1844 -----DDTENINESKKIKTED-----BEEKETDKV 1869

RESULT 10
US-10-732-923-3334
; Sequence 3334, Application US/10732923
; Publication No. US20050108791A1
; GENERAL INFORMATION:
; APPLICANT: Edgerton, Michael D
; TITLE OF INVENTION: TRANSGENIC PLANTS WITH IMPROVED PHENOTYPES
; FILE REFERENCE: 38-15(52796)C
; CURRENT APPLICATION NUMBER: US/10/732,923
; CURRENT FILING DATE: 2003-12-10
; PRIOR APPLICATION NUMBER: 10/310,154
; PRIOR FILING DATE: 2002-12-04
; NUMBER OF SEQ ID NOS: 24149
; SEQ ID NO 3334
; LENGTH: 1875
; TYPE: PXT
; ORGANISM: Saccharomyces cerevisiae
US-10-732-923-3334

Query Match 3.1%; Score 305.5; DB 17; Length 1875;
Best Local Similarity 19.6%; Pred. No. 3.5e-08;
Matches 340; Conservative 271; Mismatches 627; Indels 494; Gaps 75;

Qy 60 ANFTSDYOSVKKALNGK-----TP-----DPKSSFTDPVSKFDFL---TNNGR 101
Db 380 AKSSSDFIPLKKQIKERRKTEHLQNIETFIVELEHKVPIINSFKERTDMLNELNAA 439
Qy 102 TVLEIPKQYVVISFSPEDDKERFLGPHLKEKLEDGNIQASATKFIYLLPLDMPKAAL 161
Db 440 LLL-----HTSNKNAKVKELNNAKQKJECENDLQTLTK-----QELDLCR--- 482
Qy 162 GOYSYIVDKNFNNLIHPLSNFSAOSIKPLALTRSSDFIAKLQ-----FNNQDBLWYL 216
Db 483 -QIYL-----LITNSVNSDKGFLR---KEBIOFIQNIQWDDSDTITESDQKVT 530
Qy 217 EKFFDLEALKAMIRLOTADFSFK--GNLVPFVVSFINPQONQEWASDLNODQKTVRL 274
Db 531 ERLVPEPKNI---IQLOEKNAELLKVVRNLADKLSEKSKSQSLQKIESETNEAKEAI 587
Qy 275 YLRTEFSPQAKTILDKYKDE--TFLSSIDLKASNGT--SLPANENDLKQDLVDLLDV 330
Db 588 TLKSKNDLESRIEELQKELEULKTSVPNEDASYSNVTIKQITETKRDLESQVQDLQTRI 647
Qy 331 SDYFGQSQSETIT--SNSQVKVPASERSLDRVKFKDQKQPRIKPSLYEYDALSFYSOL 389
Db 648 SQITRESTENMSLLNKIEQLDYDSKSDISIKLGEKSSRILAEERFKL----- 695
Qy 390 QELVSKPNSIKOLVNAATLARNRFLSGKGNFIFDDLAHLDYFVFLVSKAKIKQSSITKXL 449
Db 696 -----LSNTLDITKAEN-----DQLRKFDY---LQNTILKQDSKTHET 731
Qy 450 FIELPIKISIKSILQDQEPNIKTLPEKEVTFKLDNFRDVEIEK-----APG 496

Db 732 LNEY-VSCSKLSIVETELNLKX--EQKLKVHLEKNLKOELNKLSPKDSLRIMVTQLQ 788
Qy 497 LLYPGVNEELEQARKA-QRASPEKESKKGILKEFSQOKEKSKAINNQEGLEDDNITER 555
Db 789 TLQKREDLLEETRKSCQKKIDELDALSELKETSQKDHIIKOL-----BEDN----- 837
Qy 556 LPENSPIOQOENAGLSPDKPMIKDVONQYRYLAKSQ--IQELIKAKDYTKLAKLLS 613
Db 838 ---NSNIWQ-QKIEALKKDYESVITSVDKQTDIEKLQYKVKSEKEIEEDI----- 888
Qy 614 NRHTYNI SLKQLFDVNPRIPISSRDIEKAKFVLQK--TEKNKYWQIYSSASVPFQNKW 671
Db 889 RLHTYN---MDETINDDSL---KELEKSKINLTDAYSQIKKEYKDYLTTSQLOOTN 941
Qy 672 SLFGYRYLLGLDP-----KQTHIELVKLGQAGLOFEGYENLPSPFNLEDLKNIR 722
Db 942 S-----KLDESFXDFTNQIKNLTDEKTSLEDKISLLEKEQMFNLNLEDLQ----- 986
Qy 723 IKTPLFSQKDNFK--LSLLDFNNYDGEIKAPEFGLPLFLPKELRRNSSNGSGSQNSNP 780
Db 987 -KKGMEKADFKKRISILQNNNKEVEAVKSEVESKLSKIQNDLDQOTIVANTAQNN--- 1042
Qy 781 WQOEI-----ISQFKDQ-----NLSNOQOLAQFSTKIWEKIIGENEFDQ 820
Db 1043 YEQELQKHADVSKTISELRQLHTYKQGVKTLNLS-RDQL-----EN----- 1083
Qy 821 NNRLOYKLLXDLQESMINKTRDNLWYLYGBLKLKVKPKNNLEAKFROISNLQELL---TA 877
Db 1084 -----ALKENKSWSSQKESLL-----EQDLS-----NSRIELSSQNKLLYDQIQ 1125
Qy 878 FYTSAALSNWNYQDQSGAKSTIIF-----BEIAELDPKV----- 912
Db 1126 IYTAADKEVN-NSTNGPGLANNILTLRERBILDQTKVTVRERDAKMLRQKISLMDVELQD 1184
Qy 913 -KEKGVADVQLKPHYAI--GFDNAGKFNQ-EVIRSSRTI--YLTSGSKSLEADT-I 965
Db 1185 ARTKLDNRSVEKENHSHIIQQHDDIMEKLANLNLRESNITLRNELENNNNKKELQSEL 1244
Qy 966 DOLNAQVKNAPLGL-----QSFYLDTERFGVFOKFLATSLAVQHKQKTELPK 1012
Db 1245 DKLQNV--APIESALTALKYMOEKOELKLAKEVHRWKRSODILEKHEQSSSDYE 1302
Qy 1013 KLNNDGVTLHD-----KLKPVIPQISSSPKDFEGKLNQNGSQNV 1056
Db 1303 KLESEIENLELENKERQGAEEKENRLEKQERLKT-----KLSQDSLTEQV 1354
Qy 1057 N-VSTFGSIIESPYFTNFOBADLDQDQGBDSQGNNSLD----- 1096
Db 1355 NSLRDAXNVLENSLSEAN---ARIEELQNAKVAQGNQLEAIRKLOBDAEKASRELOAK 1410
Qy 1097 -----NQEAGLLK-----OKLAILLQNO----- 1114
Db 1411 LEESTSYESTINGNEEITTLKEEIEKQRIQOOLQATSANEQNDLSNIVESMKKSPEE 1470
Qy 1115 -FIQVYQNDKEIIFEI-----IHVEKV-----SELSFRVEFKLAKTLBNGK 1156
Db 1471 DKIFEIKETQVNEKILEAQERLNQPSNIWEEIKKKWESEHEQEVSQKIREASEALKK 1530
Qy 1157 TIRVLSDTMSLIVNTTIE-----KTPEMSAPVEVPDFTKWVEQYDPTPLA 1202
Db 1531 RIRLPTERKNIIRKEEKEEKEPEEKVEBERIKSMEQSGEIDVVLRKOLEAKVOEKOKE 1590
Qy 1203 AKTKFVLKFDQI-PVDGSGNIS---DKMLASPLVHIQOMLRSLPVVKTIRELGLKTE 1257
Db 1591 LENEYKGLQBELKDXDPVPHSSHISODDERDKLAEIESRUREEFNNELOAIKK-----KSF 1644
Qy 1258 OOOOQOQOQOQOQOKKAVRKEEEL-ETYN-----PKDEFNILNPLTKAHLTLNLSNVN 1311
Db 1645 DEGKQOAMKTYLLERKLAKMESQLETKOABSPKSVNNVQPLGLPRKIBEN--SN 1702
Qy 1312 DP-NYKIEDLVKINEAGDHQAFSLRANNIKRLMNTPIITPADYNPFYFNEDWRSIDKY 1370
Db 1703 SPFNPLSGEKLKLNKSSSGGFN-----PFTSPSPKHLQNDNDKR---ES 1747

Db 1842 --EENTLALQELHENLEEMSVKMERDNLRRVEETLKL--ERDLKESLOETKARDEI 1897
 Qy 1306 SNLVNDPNYKIEDOLKVIKNBAGOH-QLAFSLRANNIKRLMNTPIITFADYNPFYINEDW 1364
 Db 1898 Q-----QELKTARMLSKHEKVTVDKL--EKISEKTIQISDIQ----- 1933
 Qy 1365 RSIDKYLANKGNVSHOQAAAGNQGSLIQLRLNKNIKPFTPTPALIALKDRNNTNLSNY 1424
 Db 1934 KDLDK---SKDELQKDRQ-----NHQVKEP-----RLLSDGOQHLMESL 1971
 Qy 1425 SDKIIMPKVILVERSIGVPMSTGLDGYIGSEOTKDGTSSSSQKGFQDQFIQALGL--- 1481
 Db 1972 REKSRKE--LLKR-----YSEMDHY-----ECINRLSLDLE 2003
 Qy 1482 KNTYHG-----KLGLSIRIFDPGNELAKIKDASNKGEELKLSYDLFKNYLNEYEKKS 1536
 Db 2004 KEIEFHRIMKKLYVLSY-----VTKIEE-----QHECINKFEM--DFIDEVEKQK 2048
 Qy 1537 PKIAGWNIHDPQKEYPNPQKLPENYLNVLNQPWKVTL-----YNSDF----- 1583
 Db 2049 ELLIK-----IQHLOQDCVPSRELAD---LKLQNMDLHIEILKOFSESEFPIKTEF 2100
 Qy 1584 ---ITN-----LFVEPEGSDRSGTKLQVIOK-----QVNNYADWGSAYLTFWYDKN 1629
 Db 2101 QOVLSNRKEMTOFLEEWLNTFRDIEKLKNGIQKENDRICOVNN-----FFNRI 2149
 Qy 1630 I-ITNPNVI---TANADVFIDKVKEL--EONTKLIAPNITQWPNISGSKEKFKYKTVF 1684
 Db 2150 IAINMESTEFERSATISKWEFODLKLKKEKELFKNYQTLKTSLSAQ---VNPT-- 2204
 Qy 1685 FGNWENENSSMNSQAOTPTWEKIRE 1709
 Db 2205 ---TODKNPHVTSRATQLTTEKIRE 2227

RESULT 14
 US-10-828-985A-7
 ; Sequence 7, Application US/10828985A
 ; Publication No. US20050003402A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Amour, Christopher D
 ; APPLICANT: Castle, John C
 ; APPLICANT: Garrett-Engle, Philip W
 ; APPLICANT: Kan, Zhengyan
 ; APPLICANT: Loerch, Patrick M
 ; APPLICANT: Tsinoornas, Nicholas F
 ; TITLE OF INVENTION: Novel Isoforms of Centromere Protein E (CENPE)
 ; FILE REFERENCE: RS0210Y
 ; CURRENT APPLICATION NUMBER: US/10/828,985A
 ; CURRENT FILING DATE: 2004-04-21
 ; PRIOR APPLICATION NUMBER: US 60/464,905
 ; PRIOR FILING DATE: 2003-04-23
 ; PRIOR APPLICATION NUMBER: US 60/510,701
 ; PRIOR FILING DATE: 2003-10-10
 ; NUMBER OF SEQ ID NOS: 25
 ; SOFTWARE: PatentIn version 3.2
 ; SEQ ID NO 7
 ; LENGTH: 2568
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 ; US-10-828-985A-7

Query Match 3.1%; Score 305.5; DB 17; Length 2568;
 Best Local Similarity 19.3%; Pred. No. 5.5e-08;
 Matches 363; Conservative 305; Mismatches 646; Indels 571; Gaps 91;

Qy 1 MNKKSTLLLATAAAIIGSTVFTVGLASKVKYRGVNTQGVISQLGIDSVAFKPSIA 60
 Db 763 IQDKSELHIITSEK---DKLFSEVVHKESSRV-----QGLLEIG-----KTKDDLA 806
 Qy 61 NPTSDYQSVKALLNGKTFDPKSSSEFTDFVSKFDFLTNNGRTVLEIPKKYQVVISFSP 120

Db 807 TTQSNYKSTQDEFQNFKTLH-----MDFEQYKVMVLEE----- 839
 Qy 121 DDKERFRLGHLKEKLEEDGNIAQASATKFIYLL-----PLDMPKAAALGOYVYVDKPNFL 175
 Db 840 --NERM-----NOEIVNLSKEAQFDSLSGALKTSLSYKTOEQLEKTRVQERLNEM 889
 Qy 176 --IIHPLSNFSAQSIKPL-ALTRSSDFIA-KLQFNQDDELWVLYLEFFDFLEALKANIRL 231
 Db 890 EQLEKEQLENRDS-----PLQTVREKTLITELQOQ--TLEEVKTLTQEKDDLKQLESQI 943
 Qy 232 QTADFSPEKGNLVDPPFYVSPFIRPNQKQWASDLNQDKTVRLYLRTFEPQAKTIL--- 288
 Db 944 ERDQL---KSDIHT-VNMMIDTQELRNALLESKQHOETINT-LKSKI SEEVSRLHME 998
 Qy 289 -KDYKYDETFLSI-----DLKASNGTSLPAN-----ENDL--- 319
 Db 999 ENTGETKDEFQKQWGDGKODLEAKNTQTLTADVKNNEIEQORKIFSLIOEKNEVLOQM 1058
 Qy 320 -----KDQLDVLDD-----VSDYFGQSETIT--SNSQVQKVPASERS 356
 Db 1059 LESVIAKEQQLKTDLKENIEMTIENQOEELLLGDELKKQOEIVAEQKHAIKKSGELSR 1118
 Qy 357 LKDRVKPKDQKQPRIEKFSLYDYDALSFYQQLVSVKPNKSIKDLVNATLARNLRLSLG 416
 Db 1119 C-DRLAEVBEKLEKSQQLQEQKQOQLLNVOESEMOKINEIENLKNE--LKNKELTLE 1175
 Qy 417 KYNFLPDDLASHLDYFLVSKAKIKQSSITKGLPIELPIKISLKSSILGDOEPNITLFE 476
 Db 1176 HMETERLELAQKLENYEEVKSITKRRKVKLKL-----QKSF 1213
 Qy 477 KEVTFKLDNFRDVEIEKAFGLLYPGVNEELEQAKAQRASPEKEKSKGLKEFSQKKEEN 536
 Db 1214 TERDHLRGYREIE---ATGL---QTKELKIAH-----IHLKEHETIDELRSVSEK 1261
 Qy 537 SKAINNOEGLEEDDNITERLPENSPIYOQENAGLQASPKPYMKDVQVQRYLAKSQI 596
 Db 1262 TQAIINTQDLKESHT--KLQEEIPVLHEEDEL-----LPNVKQVSETQETMNELELL 1311
 Qy 597 QELIKADYTKLAKLSNRHTYNTISLAKGBLFDVNPRIPS-SRDIEKAKFVLDKTEKNK 655
 Db 1312 TEQSTTKDSTTLARIEMER-----LRLNEKFSQSEIEISLTKERONMLTKIKALEV-K 1364
 Qy 656 YWQIYSSASPVFQNKWSLFGYRYLLGLDPRQTTHE-LVKLGQKAGQFQEGYENLPSPDN 714
 Db 1365 HDQL-----KEHIRETLAKIQESQSKQ-----EQSLN 1391
 Qy 715 LEDLKNIRIKTLPFSQKDNFKLSLLDFNNYVDGEIKAPFGLPLPLPKELRRNSNSGGS 774
 Db 1392 MKEKDN--ETTKIVSEMEQKPK-----DSALLRIEIEML-LGLSKRLOESHQD----- 1435
 Qy 775 QNSNSPWEQEIISQFQKQNLNSQDLAQFSTKIWE---KIIGDENEPDNNRLOYKLLKD 831
 Db 1436 -----EMKSVAKEKDDLQRLQEVLOSEDLQENIKKEIYAKHLETEEEELKVAHCCLKE 1488
 Qy 832 LOESWINKTRDNLNWTYLGDKLVKPKNNLEAKFRQISNLQELLTAFYTSAAALSNNMY 891
 Db 1489 -QETINELR-----VNLSEKETEISTIQKLEAI--NDKLQN----- 1523
 Qy 892 QDSGAKSTIIFEEIABLDPKVKVKGADYQQLKHYAIGFDNDNAGKQFQVIRSSRTIY 951
 Db 1524 -----KIQEIVEKEEQLNIKQISEVQENNVNKL-----Q 1552
 Qy 952 LKTSKSKLEADTIDOLNOAVKNAPLGLQSFVLDTFRFGVQKLATSLAVQHKOKETLP 1011
 Db 1553 FKHEHRKAK-----DSALQSIKSMLEL-----TNR--LOESQEEIQIMKEKEE--- 1594
 Qy 1012 KKLNDNGVTLIHDKLK---KPVIPQISSSPBKDW-----FE 1044
 Db 1595 MKRVEALQIERDQLKENTKEIVAKKESQSEKEYQFLKMTAVNETQKMCIEHLKEQFE 1654
 Qy 1045 -GKLN-QNGSQONVNVSTFGSIIESPYSTFQEDADLDQDQDDSRQGNISLDNQEAGL 1102
 Db 1655 TQKLANLENIETNIRLTQI-----LHENLEEMRSVTKE-RDDLRSVBETL-KVERDQ 1704


```
Db 814 IGEHKKKEVSIREEHKKSELEKSHADVAE-----LEE 860
QY 1122 NDKIEPIINV-----EKVSELSRVEFKLAKTLEONGKTIIRVLSDETMSLIVNTTIEK 1176
Db 861 QHKE-----EVVKIGEHREVEKIEEHKKEVSIREEHKKVNL-EKRHAAVTELEEH 916
QY 1177 TPMSAVEPFDTKWVEYDPRTPLAATKFKVLPKQOIPVDGSGNISDKWLASIPLVIH 1236
Db 917 KKENSELEK-----NSADVTELEKREBELGMGEHIEEMGKIEEKRESEVRIEE 969
QY 1237 QOMLELSPVKITRELGLTKSQOQOQ-----OQOQOQOPOKKAVRKEE 1281
Db 970 QHKEVALEKKGHTDVVVMLEBEQHKKEENVLEKSHADVVVLEEKREVEKSGEQHKEE 1029
QY 1282 LETYNPKDEFILNPLTKAHLRTLSNLVN-----NDPNKYIEDLK 1321
Db 1030 VAELE-KNNSALLDVEENHSEVSLKEEHKESTRELVELTSLQKNEBIIQNSNNKIEELT 1088
QY 1322 VIKNEAGHQALFASLRANNIKRLMPTPTFADYNPFYFYNEDWRSIDKYLNKNGVSSHQ 1381
Db 1089 NVIKLNDLSIMCYKKKISSEDEKRE-----YNBIEQKLNKRLNEMKDKSDE 1136
QY 1382 QQAAGNQGSLIQRLNK-----NIKPTFTPALIALKDRANTNLNSYSDKILMIKP 1433
Db 1137 I-----IKKEDIKKLNKKNISNYKIFIKENTY-----KNSKMLINENKEEIIVDN 1182
QY 1434 --KYLVERSIGVPWSTGLDGYIGSBQTKDGTSSSQKQFGDQDTQALGLKNTYHGKL- 1490
Db 1183 ICKEKISESDIEGKGNLKMILSLKKERSVFSVDNKNENNELVDI--VENT-YINKIE 1239
QY 1491 -----GLSIRIFDPGNELAKIKDASNNKGBE--KLKSYDLFPQVLYNEYEK 1534
Db 1240 VYKBEIKDNKNMEDKNKIILDSNELANLKNVKNVLTENDNLKEMEMQKNVNDEN 1299
QY 1535 -----KSPKIAKW-----TNHPDQKEVPNPNOKLPENLVNLVNOPWKVTL 1577
Db 1300 IMLNDDIIFLTKIESEWKBNKILENDIEQINKY-----KIKEENLMKINE----- 1348
QY 1578 YNSDPFITNLFVEPE-GSDRSGTGLKQVIOQVNNVADMGSAVLTTFWYDKNIITNQPN 1636
Db 1349 --SINEVNLKNQIELEKKEKNLNTSNVNNKIVENNILENG-----NLKQNSN 1396
QY 1637 VITANIADVFIKDVKELENT 1657
Db 1397 -----KEIEDKT 1403
```

RESULT 16

```
US-10-450-763-36864
; Sequence 36864, Application US/10450763
; Publication No. US20050196754A1
```

GENERAL INFORMATION:

```
; APPLICANT: Hyseq, Inc
```

```
; TITLE OF INVENTION: NOVEL NUCLEIC ACIDS AND POLYPEPTIDES
```

```
; FILE REFERENCE: 790CIP3/US
```

```
; CURRENT APPLICATION NUMBER: US/10/450,763
```

```
; CURRENT FILING DATE: 2003-06-11
```

```
; PRIOR APPLICATION NUMBER: PCT/US01/08631
```

```
; PRIOR FILING DATE: 2001-03-30
```

```
; PRIOR APPLICATION NUMBER: 09/540,217
```

```
; PRIOR FILING DATE: 2000-03-31
```

```
; PRIOR APPLICATION NUMBER: 09/649,167
```

```
; PRIOR FILING DATE: 2000-08-23
```

```
; NUMBER OF SEQ ID NOS: 60736
```

```
; SOFTWARE: Custom
```

```
; SEQ ID NO 36864
```

```
; LENGTH: 2633
```

```
; TYPE: PRT
```

```
; ORGANISM: Homo sapiens
```

```
; FEATURE:
```

```
; NAME/KEY: DOMAIN
```

```
; LOCATION: (77)..(99)
```

```
; OTHER INFORMATION: KINESIN HEAVY CHAIN SIGNATURE domain identified by eMATRIX,
; OTHER INFORMATION: accession number PR00380A, p-value=1.000e-25, raw score of 14.18
; FEATURE:
```

```
; NAME/KEY: DOMAIN
```

```
; LOCATION: (12)..(2412)
```

```
; OTHER INFORMATION: Kinesin motor domain identified by Pfam, accession name
; OTHER INFORMATION: kinesin, E-value=4.9e-217, Pfam score of 734.4
```

```
; FEATURE:
```

```
; NAME/KEY: misc feature
```

```
; LOCATION: (1)..(2633)
```

```
; OTHER INFORMATION: Xaa = X or * as defined in Table 2
```

```
US-10-450-763-36864
```

```
Query Match 3.1%; Score 303; DB 18; Length 2633;
```

```
Best Local Similarity 19.5%; Pred. No. 7.9e-08;
```

```
Matches 370; Conservative 322; Mismatches 688; Indels 520; Gaps 91;
```

```
QY 1 MNQKSTLLATAAAIIGSTVFGTVVGLASKVYRGVNVPTQGVISQIGLSDVAFKPSIA 60
```

```
Db 763 IQDKSEELHIITSEK---DKLFSEVVHKESRV-----QGLLEEIG-----KTQDDLA 806
```

```
QY 61 NFTSDYOSVKALLNGKTFDPKSEFTDFVSKFDFTLNGRTVLEIPKQYVVISEFSPE 120
```

```
Db 807 TTQSNYKSTQDEFQNFKTLH-----MDFEQKYKMWLEE----- 839
```

```
QY 121 DDKERFRLGFLHKEKLEDGNIQAQATKFIYLL-----PLDMPKAALQOYSYIVDKNFNNL 175
```

```
Db 840 --NERM-----NQEIVNLSKEAQKFDLSLGAALKTELSTYKQELQEKTRVQERLNEM 889
```

```
QY 176 --IHLPLSNFSAOSIKPL-ALTRSSDFIA-KLQNFNNQDELWVLYLEKFPDLEALKANIRL 231
```

```
Db 890 EQLKEQLENRDS---PLQTVREKTLITKELQ--TLEEVKTLTQKDDLKQLQESLQI 943
```

```
QY 232 QTADFSFEKGNLVPFVYSFIRNPQOKENASDLAQOKTVRLYLRTEFSQAATIL--- 288
```

```
Db 944 ERDQL--KSDIHT-VNMNIDTQELRNALLESKQHQETINT-LKSKISEVSRNLHME 998
```

```
QY 289 -KDYKYKDEYFLSSI-----DLKASNGTSLFAN-----ENDL--- 319
```

```
Db 999 ENTGETKDEFOQKMGVGDKKQDLKAKTQHTADVKNENIEQQRKIFSLIQEKNELOOM 1058
```

```
QY 320 -----KQQLDVLDD-----VSDYFGQSETIT--SNSQVKVPVPSASRS 356
```

```
Db 1059 LESVIAKEQLKTDLKENIEMTIEQBELRLGDELKKQEIIVAEQKNHAIKKEGELSRT 1118
```

```
QY 357 LKDRVKFKQOQKPRIEKFSLYEDALSPYLOELVSKNSIKDLVNATLARNLRFSLG 416
```

```
Db 1119 C-DRLAEVEEKLEKSKQLOEQKQOQLLVQSEMSEMQKKEINEIENLKN--LKNKELTLE 1175
```

```
QY 417 KYNFLDDELASHLDYFVLSKAKIKOSSITKFLPIELPIKISLSSILGDQEPNIKTLPE 476
```

```
Db 1176 HMETERLELAQKNENVEEVKSITKERVKLKEL-----QKSFE 1213
```

```
QY 477 KEVTFKLDNFRDVEIERKAFGLLYPGVNEELQARAKQASFEKSKKGLKEFSQOKEEN 536
```

```
Db 1214 TERDHLRCYIREIE--ATGL---QTKELKIAH-----IHLKEHQSTIDELRRSVSEK 1261
```

```
QY 537 SKAINNQGLEEDDNITERLPENSPIQYQENAGLGASPDYPMIKDVONQRYYLAKSQI 596
```

```
Db 1262 TQAINTQDLSEKSHT---KLOEEIPVLHEEQEL-----LPNVKVKVSETQETMNELELL 1311
```

```
QY 597 QELIKAKDYTKLAKLSNRHTYINSLRKEQLFDVNPRIPS-SRDIEKAKFVLDKTEKNK 655
```

```
Db 1312 TEQSTTKDSTTLARIEMER-----LRNKEFQSQOBEIKSLTKERDNLKTIKEALEV-K 1364
```

```
QY 656 YWQIYSSASPVFQNKWSLFGYYRYLLGLDLPQKOTIHE-LVKLGQKAGLQFEGYENLPSPFN 714
```

```
Db 1365 HDQL-----KEHIRETLAKIOESQSK-----EQSLN 1391
```

```
QY 715 LEDLKNIRIKTPLFSQKONFKLSLLDFNNYVDGEIKAFEPGLPLPLKELARNSSNGGS 774
```

```
Db 1392 MKEKDN--ETTKIVSEMEQFKPK-----DSALLRIEIM-LGLSKRLQESHQD----- 1435
```

```
QY 775 QNSNPWEQEIISQFQONVSNQDLAQFSTKIWE-----KIIGDENBPDQNNRLOQYKLLKD 831
Db 1436 -----EMKSAVAKEDDQRLQVLOSQDQLENKEIKETIAVXHELETEBELKVHAHCCLKE 1488
QY 832 LOESWINKTRDNL-----YWTYLGDKLVKVPKKNLEAK-----PROISNLOQELL 875
Db 1489 -QESTINELRVNLSSEKETEISTTQKQLEANDKLQNKIYIYKEBEQNLNTKIQISEVQENV 1547
QY 876 TAF-----YTSAAASNWNYYQDSGAKSTIIPBEIAELDPKVKEKVGADV 920
Db 1548 NELQKPEHRKAKDSALQIESKMLELTNRLQESQEBEIQIWIKEKEEM-KRVQEAQIER 1606
QY 921 YOLKPHAIGFDDNAGKPNQEVIRSSRTI-YLKTSGSKLEAD--TIDOLQNAVKNAPL 977
Db 1607 DOLK-----ENTKEIVAKMESQEKYQFLKMTAVNETQKMCBEIHLKESQFETQKL 1658
QY 978 GLOSFYLDTRFRFGVQKLAATSLAVQH-----KQEKXTPKKNLNDGYTLIHDKLKK 1028
Db 1659 NLEN--IETENIRLTQILHENLEMRSVTKERDDLRSVEETL--KVERD--QIKENLRE 1711
QY 1029 PVIQISSPEKOWPEGLKNGQSNVNVSTFGSIIIESPYFTNFQEDADLDQDQ--- 1084
Db 1712 TITRDLKQEBELKIVHMLKHEH---QETIDKRGIVSEKTEISNMQKOLEHSNDALKAQ 1768
QY 1085 -----QDSROGNSLNDQ-----AGLLKQKLAILL-----GNQFYQYQNDKEI 1126
Db 1769 DLKIQEELRIAHMLKEQOETIDKLRGIVSEKTDKLSNMQKOLENSNAKLQEIQELKAN 1828
QY 1127 EPEITNV-----EKVSLSFRVEPKLAKTLEDNGKTRVLSDETMSLIVNTTEKTP 1178
Db 1829 EHQUITUKKDVNETQKQSEME-----QLKKQIKQOGLTUSKLEIENLNL-AOELHENLE 1882
QY 1179 EMSAPPE-----VFDTKVEQVDPRTPLAARTKFLVKFDQIPVDGSGNISDKWLASI 1231
Db 1883 EMKSMKERDNLNRVEETLKLRLDQKLESIQETKARDLEIQEELK-----TCSEK----- 1932
QY 1232 PLVTHQOMLRLSPVVKTIIRGLKTEQOQQOQQOQQOQQOQQKAVRKEBELETYNPDEF 1291
Db 1933 ---ISEXTIQISDIQ-----DLDSKDELQKIQEIQELQKQLLRVKVEDVNNSHKK--I 1982
QY 1292 NILNPLTKAHLRLTSLNVNNDPNVKIEDLKVINEAGDHQALFSLRANNIKLMMNTPITF 1351
Db 1983 NEMEQLKKQF-----EPNV-----LCKEMDNFQLTKKLH-----ESLEETRIVA 2022
QY 1352 ADYNPFYFYNEDWRSIDKYLNNKGN--VSSHQOQAAAGNGSGLIQRLNKNIKPETFPA 1409
Db 2023 KE-----RDELRIKESLKMREDQFIATLREMIARDQ-----NHQVKEK----- 2063
QY 1410 LIALKDBRNTNLSYDKIIMIKPKYLVERSIGVPWSTGLDGYIGSEQTKDGTSSSSQK 1469
Db 2064 -RLLSGQQQLHMSLEKCSRIKE--LLKR-----YSEMDHNY----- 2098
QY 1470 GFQDQFIOALGL---KNTEYHG-----KLGLSIRIPDPGNELAKIKDASNKKEEKLKS 1521
Db 2099 -----ECLNRLSLDLEKIEFHRTMKKLKYVLSY-----VTKIEE-----QHCEINK 2141
QY 1522 YDLFKNYLNEYKSKPIAKGWTNIHPDQKEYPNPNQKLPENYLNVLNQPKWVTL----- 1577
Db 2142 FEM--DFIDEVEKQKELLIK---IQHLQDCDVPSRELRD---LKLQNMDLHIEEIL 2191
QY 1578 --YNSDP-----ITN-----LFVPEGSDRSGTKLKQVIOK-----QVNNY 1614
Db 2192 KDFSESEFPISIKTEFOQLVLSNRKEMTQFLEEWLNTRFDIEKLKNGIQKENDRICQVNN-- 2249
QY 1615 ADMGSAYLTWYDKNI--ITNQPNVI---TANIAADVFIKQVKEL-EDNTKLIAPNITQWMP 1669
Db 2250 -----FFNRIITAINWESTEFERSATISKEWEQDLKSLKKEKELFKYQYTLKTS 2300
QY 1670 NIGSKEKFKYKPTVFFGNWENENSNMSQAOTPTWEKIRE 1709
Db 2301 LASGAQ---VNPT-----TQDNKNPHVTSRATQLTTEKIRE 2333
```

```
RESULT 17
US-10-723-860-749
; Sequence 749, Application US/10723860
; Publication No. US20040253606A1
; GENERAL INFORMATION:
; APPLICANT: Aziz, Natasha
; APPLICANT: Ginsburg, Wendy M.
; APPLICANT: Zlotnik, Albert
; TITLE OF INVENTION: Methods of Diagnosis of Soft Tissue Sarcoma, Compositions &
; FILE OF INVENTION: Methods for Screening for Soft Tissue Sarcoma Modulators
; TITLE OF INVENTION: Methods for Screening for Soft Tissue Sarcoma Modulators
; FILE REFERENCE: 05882.0193 NPU01
; CURRENT APPLICATION NUMBER: US/10/723,860
; CURRENT FILING DATE: 2003-11-26
; PRIOR APPLICATION NUMBER: 60/429,739
; PRIOR FILING DATE: 2002-11-26
; NUMBER OF SEQ ID NOS: 8393
; SOFTWARE: Patent in version 3.2
; SEQ ID NO 749
; LENGTH: 2663
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-723-860-749
```

```
Query Match 3.1%; Score 301; DB 16; Length 2663;
Best Local Similarity 19.5%; Pred. No. 1.1e-07;
Matches 373; Conservative 320; Mismatches 689; Indels 530; Gaps 92;
```

```
QY 1 MNKKSTLLATAAAIIGSTVFVGLASKYKRGVNPQGVISQGLIDSVAFPKPSIA 60
Db 763 IQDKSEELHIITSEK---DKLSEVVHVESRV-----QGLLEIG-----KTDDLA 806
QY 61 NFTSYQSVYKALLNGKTPDPKSEBTFVSKPFDLTNNGRTVLEIPKPKYQVVISFSP 120
Db 807 TTQSNYKSTQEQFQNPCTLH-----MDPEQYKMWLE----- 839
QY 121 DOKERFRLGFLHKEKLEDGNIAGSATKFIYLL-----PLDMPKALGOYSYIVDKNPNL 175
Db 840 --NERM-----NOEIVNLSKEAQFDSGLALKTELSTYQELQEKTRVQERLNM 889
QY 176 --IIHPLSNFSAQSIKPL-ALTRSSDFIA-KLNQNNQDELWVYLEKFPDLEALKANIRL 231
Db 890 EQLKQLENRDS---PLQTVREKTLITEKLLQ--TLSEVKTLTQEKDDQLQESLQI 943
QY 232 QTADSFEEKGNLVDPFVYSFIRNPQNKQEWASDLNQDQKTVRLYLRTFSPQAKTIL--- 288
Db 944 ERDQL---KSDIHTD-VNMNIDTQELRNALSLKQHOETINT-LKSKISEEVSRLHME 998
QY 289 -KDYKVKQDFTLSSI-----DLKASNGTSLPAN-----ENDL--- 319
Db 999 ENTGETKDEFOQKMWGIDKKQDLEAKNTQTLTADYKDNBIIEOQRKIFSLIQEKLEQQM 1058
QY 320 -----KDQLDVLDD-----VSDYFGQSETIT--SNSQVFPVPASERS 356
Db 1059 LBSVIAEKEQLATDLKENIEMTIENQEEELRLGLDELKKQEIQAQEKHAIKKEGELSRT 1118
QY 357 LKORVKFKDQKQPRIKPSLYEYDALSFYQOLQELVSKPNSIKOLVNTATLARNLRFSLG 416
Db 1119 C-DRLAEVEEKLKESQQLQEKQOQLLVNQVHEMSEMKKINEIENLKN--LKNKELTLE 1175
QY 417 KYNFLPDDDLASHLDYVFLVSKAKIKOSSTTKKLFIELPIKISLKSSILGDQERNIKTPE 476
Db 1176 HMETERLELAQKLNENYEEVKSITKERVKIKEL-----OQSFE 1213
QY 477 KEVTFKLDNFRDVEIEKAFGLLYPGVNEBELEQAKRAQASFEKESKKGLEFSQOKEEN 536
Db 1214 TERDHLRGYIRIE---ATGL---QTKELKIAH-----IHLKEHQETIDELRRSVSEK 1261
QY 537 SKAINNQEGLEDDNITERLPENSPIOYQOENAGLQASPKPYMIKDVQONQRYLAKSQI 596
Db 1262 TAAIINTQDLEKSHT---KLOBEIPVLHBEQEL-----LPNVKVSSETQETMNELELL 1311
QY 597 QSLIKAKDVTKLAKLLSNRHTYNISLRKEOLFVDPNPIPS--SRDIEKAKFVLDDKTEKKN 655
```

Db 1312 TEQSTTKDSTTLARIEMER-----LRNKEFQBSQBEIKSLTKERDNLTKIKALEV-K 1364
Qy 656 YWQIYSSASPVFQNKWSLFGYRYLLGLDPKQTHIE-LVKLGQKAGIQFEGYNLPBDFN 714
Db 1365 HDQL-----XEHIRETLAKIQESQSK-----EQSLN 1391
Qy 715 LEDLKNIRIKTPLFSQKONFKLSLLDFNNYVDGEIKAPFGLPLFLPKELRNSNSGGS 774
Db 1392 MKEKN--ETTKIVSEMQFQPK-----DSALLRIEIM-LGLSKRLQESH- 1435
Qy 775 QNSNPWQOEIISQFQDNQDQAFSTKIWE---KIIGDENEFDQNNRLQYKLLK 831
Db 1436 -----EMKSAKEDDLQRLQEVQESQDLENKEIKVAKHLETEBELKVAHCLKE 1488
Qy 832 LQESWINKTRNL-----YTYLGDKLKVPKNNLEAK-----FRQISNLQELL 875
Db 1489 -QEETINELRVNLSEKTEISTIQKLEAINDKLQKIQEIVEKEQELNRIQISEVQENV 1547
Qy 876 TAF-----YTSALSNNNYQDSGAKSTIIFETIAELDPKVKKEVGADV 920
Db 1548 NELQKFKEHRKAKDSALQIESKMLELTRLQESQOEIIOIMIKEEM-KRQVQALQIER 1606
Qy 921 YOLKPHYAIGFDNAGKNOEVIQSSRTI-YLTKSGSKLEAD--TIDQLNQAVQNAPL 977
Db 1607 DOLK-----ENTKEIVAKKESQEKYQFLKMTAVNETQEKMCIEHLKEQFETQKL 1658
Qy 978 GLQSFYLDTERGVQFKLATSVAOH-----KQEKTLPKKLNNDGYTLIHDKLLK 1028
Db 1659 NLEN-IETENIRLQILHENLEEMSVTKERDRLRSVEETL--KVERD--OLKENLRE 1711
Qy 1029 PVIPOISSPEKDWEGKLNQNGQSNVSTFGSIIESPYSTFQDADLDQDQ--- 1084
Db 1712 TITROLEKQEBELKIVHMLKEH---QETIDKURGVISEKTNEISNQKDLHNSDALLK 1768
Qy 1085 -----QDSDQGNNSLDNQS-----AGLLKQKLAILL-----GNOFIQYQNDKEI 1126
Db 1769 DLKIQEELIAHMLKEQETIDKLRGIVSEKTKLSNMQKDLNSNAKIQEIKELKAN 1828
Qy 1127 EFETINV-----EKVSELSFRVEFKLAKTLBNGKTIKRVLSDETWSLIVNTTIETKP 1178
Db 1829 EHQLITLKQDVNETQKVSEME---QLKKQIDQSLTKLEIENLNL-AQELHENLE 1882
Qy 1179 EMSAYPEVDT-KWVEQ-----YDPTPLAATKFKVLKFDQIPIVDG 1219
Db 1883 EMKSVMKERDNLRRVEETLKERDQKESLQETKARDLEIQOELKTARMLSEKHEKTV- 1941
Qy 1220 SGNISDKLASIPLVTHQOMRLSPVVKTIIRGLKTEQOQOQOQOQOQOQKQKAVRKE 1279
Db 1942 --KLREK-----ISEKTIQISDIQK-----DLDKSDELQKKIQEIQKELQALLRVK 1986
Qy 1280 BELETYNPKDEFNINPLTKAHLRLTSLNLVNDPNYKIEDLKVIKNEAGDHQALFSLRAN 1339
Db 1987 EDVNMSHKK--INEMQLKKQF-----EPNY-----LCKCEMDNFQLTJKLH-- 2026
Qy 1340 NIKRLAMTPTPADYPPFYFYNEDWRSIDKYLNNKGN--VSSHQOQAAGNGSGGLIQLR 1397
Db 2027 --ESLEIRIVAKE-----RDELRIKESLQERDQFIATUREMIADRO----- 2069
Qy 1398 NKNIKPTPTPALIALKDRNNTNLSYSDKIMIKPKYLVERSIGVPMWSTGLDGYIGSEQ 1457
Db 2070 NHQVKEK-----RLISDGOOHLMESLREKCSRIKE--LLKR-----YSEMDDHY----- 2112
Qy 1458 TKDGTSSSSQKQFQDQFQALGL---KNTEYHG-----KLGLSIRIFDPGNEELAKIDA 1509
Db 2113 -----ECLNRLSLDLEKETEFRIMKLLKLYLSY-----VTIKKEE 2148
Qy 1510 SNKGBEKLKSYDLFPKYLNEYEKSPKAKGWTIHPDQKPEYPPNOKLPENYLNVL 1569
Db 2149 -----QHECINKFEM--DFIDEVEKQKELLIK---IQHLQOQDCVPSRELDR-----LKL 2193
Qy 1570 NQPMKVTL-----YNSDSE-----ITN-----LFVEPEGSDRGSTKQKQVIOK 1608
Db 2194 QNMNDLHIEILKDFSESEPPSKTEFQQVLSNRKEMTQFLEWLNTRFDIEKLKNGIQK 2253

Qy 1609 -----QVNNNYADMGSAYLTFWYDKNI-ITNOPNVI---TANIADVFIKDVKEL-EDNT 1657
Db 2254 ENDRICOVNN-----FFNRILIAIWNTESTEERSATISKEWEQDLKSLKEKNE 2302
Qy 1658 KLIAPNITOWMPNINISGSKFKYKPTVFFGNWENENSSMNSOAOPTPTWKEKIRE 1709
Db 2303 KLFKNYQTLKTSLASGAQ---VNPT-----TQDNKNPHVTSRATQLTTEKIRE 2347
RESULT 18
US-10-732-923-3354
; Sequence 3354, Application US/10732923
; Publication No. US20050108791A1
; GENERAL INFORMATION:
; APPLICANT: Edgerton, Michael D
; TITLE OF INVENTION: TRANSGENIC PLANTS WITH IMPROVED PHENOTYPES
; FILE REFERENCE: 38-15(52796)C
; CURRENT APPLICATION NUMBER: US/10/732,923
; CURRENT FILING DATE: 2003-12-10
; PRIOR APPLICATION NUMBER: 10/310,154
; PRIOR FILING DATE: 2002-12-04
; NUMBER OF SEQ ID NOS: 24149
; SEQ ID NO 3354
; LENGTH: 1527
; TYPE: PRT
; ORGANISM: Plasmodium yoelii yoelii
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(1527)
; OTHER INFORMATION: unsure at all Xaa locations
US-10-732-923-3354

Query Match 3.1%; Score 300; DB 17; Length 1527;
Best Local Similarity 19.4%; Pred. No. 5.3e-08;
Matches 336; Conservative 278; Mismatches 620; Indels 500; Gaps 81;
Qy 97 TNNRGTV-LEIP-----KKYQVV-----IGEF--SPEDDKERFRLGPHLKELEDGNIQA 143
Db 113 TRSARTVAEILPGKTKTKNTKNTIVQGGIISPYMSSNIDKNK-----NIPN 158
Qy 144 SATKPIYLLPLDMPKALGQYIYVDKNFNLIILHPLSNFSAQSIKIPALTRSSDFIAKL 203
Db 159 DNKKELFAFTFIPKSE-----NEENKEBREVEP--NINKQNEK-----RHDGDIPE 204
Qy 204 NQFNQDELWYLEKFFD--LEALKANIRLQATDF-----SFEKGNLVDPVYVFIRNP 255
Db 205 EEFNN-----LNKISDNVLGDKNNIIPFESNHINDKERTSLSKDNL-----FNLNE 251
Qy 256 QNOKEWASDLNQDQKTVRLYLRTFSPQAKTILKYKDETFLLSSIDLKASNGTSLPAN 315
Db 252 INKKYDSK-NDDNNLLFSENQ-----TLIPAKYNEVTEIAS 290
Qy 316 ENLDKQDLDVLDVSDYFGGQSETITSNVQKVPVA-----SERSLKORVKFKDQOKPRI 372
Db 291 RTP-----NYEINRVNVEINKSGIHPNIPOFNPAQLPTEINIKNNINY--DNISPKI 343
Qy 373 EKFSIYE-----YDALSFYSLQDELVSQKPS---IKDLVNATLARN--LRFS 414
Db 344 NSKGVIENLENKKEIKRISMDNFENFNPIENHINKRNTTIVRSVFPQLEYNNVENE 403
Qy 415 LGKYNFL--PDDLASHLDYVFLVSKA--KI-----KQSSI-TKKLPIELPIKIS 458
Db 404 IRKFTIAGNNTYAGIRENDIILNISPVKIEKINRRNTHQLFWNESSINSTDKLIHKVTSPP 463
Qy 459 LKSSILGQBSQENIKTLPEKVTFFKLDNFRDVE-IEKAFGLLYPGVNEBLEQAKQAR-AS 516
Db 464 LFDNINNSNTNV-----SNEIPEKGGKMEKVEKMEKM-----EKMEKMEKVEKME 510
Qy 517 FBK-EKSKKGLKEFSQ-----QKEEN-----SKAIKNOQGLE-EDDNI 552
Db 511 FEKMEKSNHVSFMSFNLDKGVSVESTSNTIAMDENFIILNDSNTINKNNIIDAESKL 570


```
QY 740 DFNN-----YYDGEIKAFEPGLPLFLPK-----ELRRNSNSG 772
Db 637 ILKNVELKHNIHVPNSYKQENKQBPYYLILVKKEIDKLVFMPKVESLINEEKNKITEG 696
QY 773 GSQNSNSPWEQEIISQF-----789
Db 697 QSDNSEPSTEGEITGQATTKPGQAGSALGDSVQAQAQEQKQAPPPVPVPEAKAQP 756
QY 790 -----KQONLSNODLAQFSTKIWE-----KIIGDENEFDONNRLOQYKLLKD 831
Db 757 TPPAPVNNKTENVSKLDYL-----EKLYEFLNTSYCHKYLIVSHSTWNEKILQYKITKE 812
QY 832 LQESWINKTRDNLWYTYLGDGLKLVKPKNNLEAKFRQI-----SNLQELLTAFTYSAALSNN 887
Db 813 -EESKLSGC-DPL-----DLLFNQNNIPVMYSMFDLSNLSQLFMEIYEKEMVCNL 863
QY 888 WNYQDSGAKSTIIFEBIAELDPKVKGVADVQLK-----924
Db 864 YKLKNDKIKN--LLEAKKVSTSVKTLSSSMQPLSLTPQDKPEVSANDTSHSTNLNN 921
QY 925 ---FHYAIGFDDNAGFNQEVIRSSRTIYLTSGKS-----KLEADTIDOL 968
Db 922 SLKLFENILSGKNKIYQELIGOKSSNFYEKILKDSDFYNESFTNFVKSADINSL 981
QY 969 NQAVNAPLQSLQSYLDTFRGFQKLAISLAVOHQK- EKTLPKLNNDGYTLIHDK 1027
Db 982 NDESKRKL-----EEDINKLKTQLSLFDLYNKYLKLERLPDKKKTGVKYN--QIK 1033
QY 1028 KPVIPQISSPEKDMFEGKLNQCSQNV--NVSTFGSIIESYFSTNFDADLDDGQ 1085
Db 1034 KLTU-----LKEQLESKLNLSNPKHVLQNFVSF-----FNKKKEAEI 1071
QY 1086 DDSRQGNLSLDNQBAGLLKOKLAILGNQFIQYQQND--KEIEFEIINVEK--VSELS 1140
Db 1072 ---AETENTLENTKI-LLKHVKGL-----VKYNGESSPLKTLSESIQTEDNYASLEN 1121
QY 1141 FRVEFKLAKLEON-----GKTIRVLSDETMSLI-----VNTT 1173
Db 1122 FKVLKLEGLKDLNLEKKKLSYSSGLHHLIAELKEVLKNKYTGNSPSENNTPVNA 1181
QY 1174 IEK-----TPMSAVPEYF-----DTKWVEQYDRTTP-----LAAKTFVLKFDQO 1214
Db 1182 LESYKFLPEGTDVATVVSSESDT--LEQSOPKKPASTHVGAESNTITTSQNVDEVD 1239
QY 1215 ---IPVDGSGNISDKWLASI-----PLVI-----HQWMLRLSPVVKTIRELGLK 1255
Db 1240 VIIVPIFGESEEDYDGLQGVVTGEAVTPSVIDNLSKIENEYEVLYLKPLAGVYRSL 1296
QY 1256 TEQOQOQOQOQOQOQOQKAVRKEELET-YNPKDEF-NIL-NPLTKAHLRTLSNLVND 1312
Db 1297 -----KKQLENNVMTFNVNKVDILNSRFNKNFNKLVESDLIPYKDLTSSNYVKD 1348
QY 1313 PNYKIEDLVKIKNEAGDHLAFLSRANNIKRLMNTPTITFADYNPFYF-----YNEDWRS 1366
Db 1349 P-----YKFLNKEKRD---KFLSSYNIKDSIDTDINFANDVLGYIKLSEYKSKDLS 1399
QY 1367 IDKYLNNK 1374
Db 1400 IKKYNDK 1407
```

RESULT 23

```
US-10-087-464-10
; Sequence 10, Application US/10087464
; Publication No. US20030059436A1
; GENERAL INFORMATION:
; APPLICANT: Chiehti, Athar
; APPLICANT: Oh, Steven
; APPLICANT: Liu, David
; APPLICANT: Goel, Vikas
; APPLICANT: Li, Xuerong
; TITLE OF INVENTION: Band 3 Antigenic Peptides, Malaria Polypeptides and Uses Thereof
```

```
; FILE REFERENCE: S1237/7019
; CURRENT APPLICATION NUMBER: US/10/087,464
; CURRENT FILING DATE: 2002-03-01
; PRIOR APPLICATION NUMBER: US 06/272,930
; PRIOR FILING DATE: 2001-03-02
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: Patentin version 3.0
; SEQ ID NO 10
; LENGTH: 1639
; TYPE: PRT
; ORGANISM: Plasmodium falciparum
; US-10-087-464-10
```

Query Match 3.0%; Score 295; DB 14; Length 1639;
Best Local Similarity 19.5%; Pred. No. 1.1e-07;
Matches 282; Conservative 204; Mismatches 478; Indels 484; Gaps 68;

```
QY 203 LNQFNQDDELWVLEKFFDLEALKANIRLOTADFESEKGNLVDPFVYSFIRPNQOKEWA 262
Db 168 IDGYEEINELLYKLNPFYDLLRAKLN-----DVCANDYCOIPFNKIRA 211
QY 263 SOLNODQKTVRLYLRTFESPOAKTILKYKYKDETFLSSIDLKASNGTSLFANENDLKQ 322
Db 212 NELDVLKLV-----FGYRKPL-----DNIKDN 234
QY 323 LDVDDLDSYFGGSETITSNSQVKVPAPASRSRKDRVKFKDQOKPRIEK-----FSLY 378
Db 235 VG-----KMEDYIKONKKTITENINEL--IEESKTTIDKNKNATKEEEKKLYQAQYDLSIY 288
QY 379 -----EYDALSFYSQLOELVSKPSIKDLVNAV-----406
Db 289 NKQLEAHNLISVLEKRIDTLKKNENIKELDKINEIKNPPPPANSNTNTLLDKNKKIE 348
QY 407 -----LARNLRSLGKYNEFLDASHLDYVYFLVSKAKIKQSSITKKLFIELPKIS 458
Db 349 EHEKEIKEIAKTIKENI--DSLFTD-PLELEYVL--REKNKV-----IDISAKVE 393
QY 459 LKSSILGDOEPNKTILFEKEVTFKLDNFRDVEIEKAFGLLYPGVNEELSOARK-----AOR 514
Db 394 TKESTEPNEYPN-----GVTYPLS-----YNDINNALNELNSFGDLINP 432
QY 515 ASFEKEKSKGLKEFQOKEENSKAINNOGLEEDDNITERLPENSPIOYQOENAGLGAS 574
Db 433 FDTKEPSKNITVD-----NERKKFINE--IKEKIKIEKKIESDKKSYE-----475
QY 575 PDKPYMIKDVQORYYLAQSOIELIKAK-----DYTKLAKLLSNRHTYINISLRKEQLF 629
Db 476 -DRSKSLNDITKEYEKL-----LNEIYDSKFNNNIDLTNFEKMMGKRYKYV-----EKL 525
QY 630 DVNPRISSRDIKAKFVLDK-TEKNKYWQIYSSASVPFQNKWSLFGYRYRYLLGL--DPK 686
Db 526 HHN-----TPASYENSKHNLEKLTAKLYMEDYSLRNIVVEKE--LKYKKNLISKIENI 578
QY 687 QTHLVLKQAGAGLQFEGY-----ENLPSDPLE--DLKNIRIKTLPFSQK-DNFKLSLL 739
Db 579 ETLVENIKODESQ--LFEKKITKDNKPKDEKILEVSDIVKVQVQKVLNKNIDELKKTQL 636
QY 740 DFNN-----YYDGEIKAPEFGLPLFLPK-----ELRRNSNSG 772
Db 637 ILKNVELKHNIHVPNSYKQENKQBPYYLILVKKEIDKLVFMPKVESLINEEKNKITEG 696
QY 773 GSQNSNSPWEQEIISQF-----789
Db 697 QSDNSEPSTEGEITGQATTKPGQAGSALGDSVQAQAQEQKQAPPPVPVPEAKAQP 756
QY 790 -----KQONLSNODLAQFSTKIWE-----KIIGDENEFDONNRLOQYKLLKD 831
Db 757 TPPAPVNNKTENVSKLDYL-----EKLYEFLNTSYCHKYLIVSHSTWNEKILQYKITKE 812
QY 832 LQESWINKTRDNLWYTYLGDGLKLVKPKNNLEAKFRQI-----SNLQELLTAFTYSAALSNN 887
Db 813 -EESKLSGC-DPL-----DLLFNQNNIPVMYSMFDLSNLSQLFMEIYEKEMVCNL 863
```


181 QY -----SNFSAQSIKPLALTRSSDFAKLNQNNODELMVYLEKFDLEALKANRLQTAD 235
 479 Db DIAGFEIPDINSFEOLCINYTH---KIQDFNFHHSF---ILEQ---SEYLRENINWEPID 530
 236 QY SFPEKGNLVD-----PFVYSFIRPNQKQWASDNLQDQKTVRLYLRTBSPQAKTILK 289
 531 Db FGQDLQPTDILIEYKQPMGIIKLDEECIMPKSDASPMELSKNFTNTH----- 580
 290 QY DYKNDFTLSSIDLKASNGTSLFANENDLKQDLD-----VDLLD-----VSDYF--- 334
 591 Db KKPSKFNKGFIIHHYAGVYENWNLQKNTDPTISESILNLLPDSQNELIADMFIND 639
 335 QY -----GQSETITNSQKVPASERSLKURVFKKQDQK-----PRIEK 374
 640 Db PHINRPOTNGG-----GNSKLK--TASQKH-KQDLKTLMDQLESTEPHFVRCILPNLEK 690
 375 QY FSLVEYDALSFYSQLOELVSKPNSIKDLVNATLARNLRFSLGKY--NFLPDDLASHLDYY 432
 691 Db RANKFDKQNLVQGR-----CNGVL-EGIRITRAGYPMRMFDEF---IQRY 733
 433 QY FLVSKAKIKQSSITKFLFELPIKISLSSILGQDEPNI---KTLFEKEVTFKLDNPRDV 489
 734 Db SIICDNEL--SSPQNTNCE---IILKFKVKNPEDFKVGLTKIFFKNGILGKLEIIRDL 787
 490 QY EIEKAFGLLYPGVN-----EELEQAKAQRASFEKEKSKG--LKFSQOK- 533
 788 Db ALKNFTDLQKVRGNLTRLVLKQIKBIQSAQIISRTWVTLDEIKSNPWWLFFHVHP 847
 534 QY -----EENSKAINN--QEGLE-----EDDNITERL-PENSPLOYQOEN-----AGL 571
 848 Db LLEDSAKVLDSKQLENQTLTVLKQSEKUTKGLTNEKREQNNQMLQDEMINTISIA 907
 572 QY GASDPKPYMIKDVQNRQYVLAQSQIQL-IRAKDYTKLAKLSNRHTYINISRLKEQLFD 630
 908 Db KEKDKLQSLRLTENK---SKHRIETLEIKLQKPEKQNDLINEHE-----KLTRRSLE 958
 631 QY VNPRIPSSRDEKAFVLDKTEKNKYQWYISASPVFQNKWSLFGYYRYLGLDPKQTIH 690
 959 Db LKHNKAV---DKHNKAV---EELNSLNKLHDSAQSELSNL-----KQOLK 992
 691 QY ELVLKQAGLQFEGEVNLPSPFNLEDLK-----NIRIKPLPFSQKD----- 732
 993 Db EITKLNQAHSEKLLPKQK-HDKSIEDNEKVSRLNENLKKLQADLNSKDVISEHGLK 1051
 733 QY -NFKLSLLDFNNYDGEIKAPE---FGLPLFLPKELRRNSNGSGSNGSNPWEQEI--- 785
 1052 Db QNEIMKELCSDYERKQKDSORTIDSILQIKKD-----EDYKSYESKIBEA 1100
 786 QY -----ISQKQDNLSNQDLAQSFSTKIWEKIIGDENEFQNNRLOKYL 828
 1101 Db REKVTLKGVKBEKSKQEIIDQYKSEIKKLKAEALASSNTKI-----TQDQSTK-ELSS 1151
 829 QY LKDLQESW---INKTRDNLVYTWLGDKLKVPKNLEAKFR---QISNLQELLTAFTVS 881
 1152 Db LKSNESKYVAIESAKQQL-----SKKIEYSSTEADYNRLOTELQKMQKQNTYEISTR 1204
 882 QY -AALSNNMNYQDSGAKSTI-----IFEEIAELDPKVKVKGVADVYOLKPHYAIG 930
 1205 Db ITELSTKLNVQTEAKSKEIENENQPPNPQFMEETHMKLVNEQT-ASLRKEKFNK-K 1262
 931 QY PFDNAGKFNQEVIRSSRTIYKTS-----GKSKLEADTIQDLQNVQVNAPLQOSFVL 984
 1263 Db LSEELQMLKERIMWGLSTMDLTPKRRSLAIGDKSMITVTVDSPFNKLENLKFOLQ----- 1318
 985 QY DTERFVQFKLATSIAVQHKQKTKLPKLNNDGYTLIHDKLKPVIPQISSPEKDWFE 1044
 1319 Db --OEQGNFOR-AYENYAI---ELQKLNKLTLTTRGLNTNTD-----YE 1354
 1045 QY GKLNQNGQSONVNVSTFGSITIEPSYFTNFQDADLDQDQD-DSRQ-----GNSLND 1097
 1355 Db KXNDQSKRITQLETKIGRLLAN---SGGDYEDHLLDDDDWESSRSGSGSGSGSTSSS 1411

1098 QY QEAGLLKOKLAILLGN-----OFIOVYQOQDKEIEF--EIIINVKVSLSFRV-----E 1144
 1412 Db ARNSLVKSESITAFHNWRGVSQDYIYQIDYITKLSSTREELANGSKTEILRLKALLRESE 1471
 1145 QY FLAKLTLEDNGKT-IRVLSDETMSLIV--NTTIKTEPMSAVPEVFDTKVQVQDPRTPPL 1201
 1472 Db DELVQVQENYKTSVHDYEQDLAQLKVGHETLLSRNKNIDINESLEIV-KGSDEYVKKLEL 1530
 1202 QY AAKTKFVLK-FKQIIPVDSGNSIDKWLASIPLVIHQOMLRLSPVVKVTIRELGLKTRQOQ 1260
 1531 Db AESAIAISKHBEQATKEMKESRSQ-----LLLVREELRTTOIL--IKDFRIKVENLE 1581
 1261 QY QOQOQOQOQOQKAVRKEBELETYNPKDEFNINL-----PLTKAHLRTLSNLVNDPNYK 1316
 1582 Db ATIEEKHQLDANK-----BEIKQIQDKLVNHLKNFENKELNEKLKEEIKNL-NRDLDFK 1635
 1317 QY IE-DLKVIKNEAGDHLAFS---LRANNI-----KRLMNTPIITADYNPFYYN 1361
 1636 Db TDITETLIK-ENKGLQLDYEDVLLVKNLQNEVEELILQBEKLQ----- 1679
 1362 QY EDWRSIDKYLNNKGVSSHQOQAAGNGSGSLQRLNKNIKPETFTPALIALKDRNNTNL 1421
 1680 Db KIDELTNNNRQLEN-----EKLNERKIVNCTKQISGLKE----- 1714
 1422 QY SNYSDKIIMIKPKYLVRSIGVPWSTGLDGYIGSEOTK-DGTSSSQKGFDDQFIQ--- 1477
 1715 Db -LVDEISIERDKLLKDKET-----LQNDLQMTWTKFDATTTTELQAQHGELDFLKXHL 1765
 1478 QY -ALGKNTYHKGKGLSIRIFDP---GNELAKIKDAS---NKKGEKLLKSYDL- 1524
 1766 Db ENQREDSEAIKTELNQSKMSTSPDIRQOKLRNELLVTKEENFSLVTKNELNLKVSdle 1825
 1525 QY FNYVLE-----YEKSKPIAKGWTNIHPDQKEYPNQKLPENYLNL---VLNQPKWVTL 1577
 1826 Db EKLVSNEQLKYWESKVDTLKALDGNALNEKHEADKTIKNLQORSIKQLEIRVENESQLSKR 1885
 1578 QY YNSDDF 1583
 1886 Db YNDENF 1891

RESULT 28
 US-10-205-823-419
 ; Sequence 419 Application US/10205823
 ; Publication No. US20030108963A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Schlegel, Robert
 ; APPLICANT: Monahan, John E.
 ; APPLICANT: Endege, Wilson O.
 ; APPLICANT: Gannavarapu, Manjula
 ; APPLICANT: Gorbacheva, Bella
 ; APPLICANT: Hoersch, Sebastian
 ; APPLICANT: Kamatkar, Shubhangi
 ; APPLICANT: Womsey, Angela M.
 ; APPLICANT: Glatt, Karen
 ; APPLICANT: Zhao, Xumei
 ; APPLICANT: Anderson, Dustin
 ; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND
 ; TITLE OF INVENTION: METHODS FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND
 ; TITLE OF INVENTION: THERAPY OF PROSTATE CANCER
 ; FILE REFERENCE: MRI-044
 ; CURRENT APPLICATION NUMBER: US/10/205,823
 ; CURRENT FILING DATE: 2002-07-25
 ; PRIOR APPLICATION NUMBER: 60/307,982
 ; PRIOR FILING DATE: 2001-07-25
 ; PRIOR APPLICATION NUMBER: 60/314,356
 ; PRIOR FILING DATE: 2001-08-22
 ; PRIOR APPLICATION NUMBER: 60/325,020
 ; PRIOR FILING DATE: 2001-09-25
 ; PRIOR APPLICATION NUMBER: 60/341,746
 ; PRIOR FILING DATE: 2001-12-12
 ; PRIOR APPLICATION NUMBER: 60/362,158
 ; PRIOR FILING DATE: 2002-03-05

```
/ NUMBER OF SEQ ID NOS: 455
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 419
; LENGTH: 1979
; TYPE: PRF
; ORGANISM: Homo sapiens
US-10-205-823-419

Query Match      2.9%; Score 286.5; DB 14; Length 1979;
Best Local Similarity 19.0%; Pred. No. 4.6e-07;
Matches 373; Conservative 345; Mismatches 682; Indels 563; Gaps 99;

Qy 22 FGTWGLASKVGRVGNPTGCVISQLGLDSVAFKPSIANFTSD-----VQSXKAL 73
||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 164 FGDIISSQBIN-RLSNEVSRLESEVGHWRHIA-QTSKAQGTDSQSEICKLQNIKEL 221
||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 74 LNKGTDPKSSSEFTDFVSKFELTN-NGRTVLPIPKKYQVVISFSPEDDKERFLGPHL 132
||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 222 KQN-----RSQEDDDHQHEMSVQLNAHQKQLTISRHRREELSDY-----EERI 265
||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 133 KEKLEDGNIAQSATKFTYLLPLDMPKAALQGYSYIVDKNFNNLIHPLSNFSAQSIPLA 192
||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 266 -EEL--NLLQOGSGV--TETLSKI-----YEMQKTIQVLQIEKVE--STKMEQLE 312
||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 193 LTRSSDFIAKLNQFNN-----QDELWVYLEKFPFDEALKANIRIQ-----TA 234
||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 313 -DKIKDINKKLSAENDRDILRREQEOLNV--EKQIMEBCE-NLKECSKLOPSAVKQS 368
||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 235 DPFSEKGNLV-----DPF-VYSFTRNPONQEMASDLNOD---QKTVRLVLRTEFSP 282
||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 369 DTWTEKERIIAQASVEEVRLOALQDAENIWRSLSLNQDNSLAEDNLKQRIEVLUE 428
||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 283 QAKTILDKYKQDETFLSSIDLK-----ASNGTSLPANENDLKOQDVLDDVSDY 333
||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 429 KEKSLLS--QEKELOQLSLKLNNEVEVIKSTATRIDSLDSELHDLRLNLE----- 477
||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 334 FGGQSETITSNQVKVPVAPASERSLKDVKPKQOQKPRIEKSFLSYEYDALSFYSQIQELV 393
||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 478 -----AKEQELNQSIS--EKETLIAEIBELDRQOEATKHMILIKDQL 518
||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 394 SKPNSIKDLVATLARNLRSGLKYNPLFD---DLASHLDYVFLVSKAKIKOSSITKULP 450
||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 519 SKQONEGDSIISKUKQDDEKRRVHQLEDDKMDITKELD-----VQKEKTIQSEVA--- 570
||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 451 IELPIKISLSGILGQEPNIKTFLFEKVTFLDNFRDVEIERKAFGLLYPGVNEELEQAR 510
||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 571 -----LNDLH-----LTQKLEDKVENL-----VDQLN 593
||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 511 KAQRASFEKSKKGLKEPFSQOKEENSKAINNQ--EGLEEDN-----ITR----- 555
||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 594 KSOENSVSIQKENLELKEHTRONEEELSRIRNELMQSLNQDSNPNKDTILLKEREAEVRN 653
||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 556 LPEN-SPIOYQOENAGIGASDPKPYMKDVONQRYILA---KSOIQELIKADYTKLAK 610
||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 654 LKQNLSELEQNLNKKVAFDVK-----MENSEKVLACEDVRHQLEECIAGNQLSLEK 707
||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 611 LLSNRHTNYSILRLKEOLFVNPRIPSSRDIKAKFVLDKTEKNKYQWYSSASPVFQNK 670
||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 708 -----NTIVETLXMEKEIEBAELCWAKKRLLEA-----NKYEKTIELSNA-RN- 751
||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 671 WSLFGYRYRLLGDPKQTIHE-LVKQKAGLOF-----EGYENLPDSF-----NLE 716
||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 752 -----LNTSALQLEHEHLIKLNQKQDMETIAELKKNIEQMDTDHKETDVLSSSLE 801
||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 717 DLKNIRIKTLPQKONKFLSLDFNNYVDGEIKAPEFGILPLPKELRN-----767
||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 802 EQQL-----TOLINKKEIIFIEKLEKRSKLOEELDK-----YSQALRKNELROTIEE 850
||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 768 SSSNGSGSQNSPWEQIIISQFKDQ-----NLSNQDLA-----QFST----- 805
||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 851 KDRSLGSMKEENHLQEELERLEEQSRTAPADPKTLDVSTELASEVSLNTIKEHLES 910
||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 806 --KIWEKIIQDENFQNNRLQYKLLKDLQESINKTRDNLNLYWYLGDKLVKPKNLEA 863
||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
```

RESULT 29

Mon Nov 21 16:44:21 2005

US-11-051-454-419
; Sequence 419, Application US/11051454
; Publication No. US20050191673A1
; GENERAL INFORMATION:
; APPLICANT: Schlegel, Robert
; APPLICANT: Monahan, John E.
; APPLICANT: Endege, Wilson O.
; APPLICANT: Gannavarapu, Manjula
; APPLICANT: Gorbacheva, Bella
; APPLICANT: Hoersch, Sebastian
; APPLICANT: Kamatkar, Shubhangi
; APPLICANT: Wasey, Angela M.
; APPLICANT: Glatt, Karen
; APPLICANT: Zhao, Xumei
; APPLICANT: Anderson, Dustin
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND
; TITLE OF INVENTION: METHODS FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND
; TITLE OF INVENTION: THERAPY OF PROSTATE CANCER
; FILE REFERENCE: MRI-044
; CURRENT APPLICATION NUMBER: US/11/051,454
; CURRENT FILING DATE: 2005-02-04
; PRIOR APPLICATION NUMBER: US/10/205,823
; PRIOR FILING DATE: 2002-07-25
; PRIOR APPLICATION NUMBER: 60/307,982
; PRIOR FILING DATE: 2001-07-25
; PRIOR APPLICATION NUMBER: 60/314,356
; PRIOR FILING DATE: 2001-08-22
; PRIOR APPLICATION NUMBER: 60/325,020
; PRIOR FILING DATE: 2001-09-25
; PRIOR APPLICATION NUMBER: 60/341,746
; PRIOR FILING DATE: 2001-12-12
; PRIOR APPLICATION NUMBER: 60/362,158
; PRIOR FILING DATE: 2002-03-05
; NUMBER OF SEQ ID NOS: 455
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 419
; LENGTH: 1979
; TYPE: PRT
; ORGANISM: Homo sapiens
US-11-051-454-419

Query Match 2.9%; Score 286.5; DB 20; Length 1979;
Best Local Similarity 19.0%; Pred. No. 4.6e-07;
Matches 373; Conservative 345; Mismatches 682; Indels 563; Gaps 99;

QY 22 FGTWGLASKVYRGVNPQTQVISQLGLDSDVAFKPSIANFTSD-----YQSVKAL 73
DB 164 FGDIISSQGEIN-RLSNEVSRLESEVGHWRHIA-QTSKAQGTDSQSEICKLQNIKEL 221
QY 74 LMGKTFDPKSSBFTDFVSKFDELTN-NGRTVLEIPKKYQVVISSEFSPEDDKERFRLGFHL 132
DB 222 KQN-----RSQETDHDQHEMSVLQNAHQOKLTEISRHRRELSDY-----EERI----- 265
QY 133 KEKLEDGNTAQSATFIYLLPDMPKAALQYSYIVDKNFNNLIITHPLSNFSAQSIKPLA 192
DB 266 -BELE--NLUQGGSGV--IETDLSKI-----YEMQKTIQVLOIEKVE--STKKWEQLE 312
QY 193 LTRSSDFIAKLQFN--OQELWVYLEKFFDLEALKANIRLQ-----TA 234
DB 313 -DKIKDINKLSAENDRIILRREQQLNV--EKQIMEECE-NLKLECSKLQPSAVKQS 368
QY 235 DFSFEKGNLV-----DPF-VYSFIRNPQONKEWASDLNOD-----QKTVRLYLATERSP 282
DB 369 DWTTEKERILAOASVEEVRLOQALSDAENIMRLSSLNQDSLAEDNLKLMRIEVL 428
QY 283 QAKTILKDVKYKDETFLSSIDLK-----ASNGTSLFANENDLKOOLDVLLDVSDY 333
DB 429 KEKSLIS--QEKELQMSLLKLNNEVEVIKSTATROISLDSSELHDLRLNLE----- 477
QY 334 FGQSQETITNSQVKVPASERSLKDVRPKDQKQKPRIEKFSLEYDALSFSYQLQELV 393
DB 478 -----AKEQELNQSIS-EKETTLAIEELDRQNOEATKMILIKOQL 518

QY 394 SKPNSIKDLVNAIARNLRFSLGKNYFLFD--DLASHLDYYPFLVSKAKIKOSSITKKLF 450
DB 519 SKQONEGDSIIISKLQDLANDEKRVHQLEDKWDITKELD-----VQEKLIQSEVA----- 570
QY 451 IELPIKISLKSSILGDOEPNITKLPBKVTFKLDNFRDVEIEKAFGLLPGVNEELEEQR 510
DB 571 -----LNDLH-----LTQKLEDKVENL-----VDQLN 593
QY 511 KAQRASPEKEKSGKGLKEFSQOKEENSKAINNO--EGLEEDDN-----ITER----- 555
DB 594 KSQESNVISQKENLELKEHITRQNEEBELSRIRNELMSLNQDSNSNPKDITLLKEREAEVRN 653
QY 556 LPEN-SPIOYQOENAGLGASDPKPYMKDVQNYRYLA-----KSQIQLIKAKDYTKLAK 610
DB 654 LKQNLSELEQLNENLKKVAFDVK-----MENEKVLACEDVRHQLECECLAGNQLSLEK 707
QY 611 LLSNRHTNYSILKLEQOLFVNPRIPSSRDIKAKFVLDKTEKNKYQIYSSASPQNK 670
DB 708 -----NTIVETLWKEKEIEAELCWAKRILLEA-----NKYEKTIIEELSNA-RN- 751
QY 671 WSLFGYRYLLGLDPKQTIHE-LVTLGQKAGLOF-----EGVENLPSDF-----NLE 716
DB 752 -----LNTSALQLEHEHLIKLNQKQDMEIAELKKNIEQMDTHUKETKDVLSLSLE 801
QY 717 DLKNIRIKTFLFSOKNFKLSLLDFNNYDGEIKAFEGFLPLFLPKELRRN----- 767
DB 802 EQKOL-----TQLINKKEIFIEKXERSKLOELDK-----YSQALRKNELRQTIEE 850
QY 768 SSNSGGSONSNPWEQIISQFKDQ-----NLSNQDLA-----QEST----- 805
DB 851 KDRSLGSMKEENNHLQEBELRREEQSRTPADPKTLDVTELASVSLNTIKEHLEE 910
QY 806 --KIWEKIIIGDENEFQDNRLQYKLLDQESMINKTRDNLTYTLGDKLVKPKNNLEA 863
DB 911 EIRHQKIIEDQO-----SKMQ--LLQSLQEQ--KKEMDE--FRYQHEQMNATHQTFL 960
QY 864 KFRQISNLQELLTAFYT-----SAALSNNWNYQDSGAKSTI 900
DB 961 KDEEIKSLQKTIQIKTLQHEERODIQTNDNDFQETKVQSLNTEENGSEKHDLSKAETER 1020
QY 901 IFEBIAELDPKV--EK--VGADVYQLKHYAIGFDDNAGKFNQEV----- 942
DB 1021 LVKGIKERELEIKLNEKNISLTQIDQLS-----KDEVGKLTQIIQKDLFIQALHA 1073
QY 943 -IRSSST--IYLTKSGSKLEADTIDQ-----LNAQVKNAPLGLQSFYLDYRFGV 992
DB 1074 RISSTSHQDVVYL-----QOOLQAYAMEREKVFVAVLNEKTR-----ENSHLKTETHKQM 1123
QY 993 QKLATSLA--VQHQKKEKTLPKLNNQDGYTLIHDKLKXPVIPQISSPEKDMFEGKLNQ 1050
DB 1124 DIVAAKEAALIKLODENKKLSTRFESSQDMFRETIQ--LSRIIR--EKDIEIDALSQK 1179
QY 1051 QGS-----QNVNVTFGSIIIESPYSTNFQDADLD-----QDQDSDS 1088
DB 1180 CQTLAVLQTSSTGNEAGVNSHOFELLOERDKLKOQVKQMEWKKQOVMTVQNMOMES 1239
QY 1089 ROGNNSLDNQAGLL--KQKLAILLGNQFIOVYQNDKEI----- 1126
DB 1240 AQLOEELHQLQVLVDSDNNNSKLQVDY-TGLIQSYEQNETKLNFGQELAQVQHSIGQL 1298
QY 1127 -----EPEIIN-----VEKVSSELSFRVFKLAKTUED 1153
DB 1299 CNTKDLGLGLDIIISPOLSSASLLTPQSAECLRASKEVLSSESSELLQOELERKLSQE 1358
QY 1154 NGKTRIVLSDE-----TMSLIWNTTIEKTPEMSAPVEFDTKWVEQYDPTPLAAKTVL 1209
DB 1359 KDATIRTLQENNRHLSDSIAATSELERKEHQDTSEIKQLK--EKQDVLQKLLKEKDL 1416
QY 1210 KFKQIIPVDGSGNISDKW-----LASIPLVHQOMLR-----SPVVKTIIR 1253
DB 1417 KAKSQDLSNENFTNKNVNEELLRQAVTNLKERILLIEMDIGKLGKGENEKIVEYR--G 1474
QY 1254 LKTEQQQQQQQQQQQQQPOPKKAVRKEEBELETYNPKDEFNINPLTKAHLTLNLVNDP 1313

Db 1475 KETEQALQETNMKFSM-----MLREKEFECHSMKERALFAEQLLKEQKGTGELNQLL 1529
Qy 1314 N--YKIEDLVKINEAGDHLAFSLRANNTIKRLMNTPTITADYNPPFYNNEDMRSIDKYL 1371
Db 1530 NAVKSMQEKTVVFOERD-QVMLAKQ---KQMENTAL-----QNEVQRLRDK-- 1573
Qy 1372 NNKGNVSSHQQAGGQSGSLIQRNLKNI--KPETFPALIALKORNTNLTNYSIDKII 1429
Db 1574 ----EPRSNQE-----LERLANHLESEDSYTRREALAEDRE-----AKLRKVT 1614
Qy 1430 MIPKXYLVERSIGVPWGTGLDYGISQTKDGTSSSQKGFQDFDIQ-----A 1478
Db 1615 VLEEK-LVSSNAME-NASHQASQVSESLOEQNLNVVSKOR--DETALQLSVSEQVKQYA 1670
Qy 1479 LGLKNTYHVGKLGSLIRIFDP-----GNELAKIKO--ASNKGEELKLSYDLFNKYL 1530
Db 1671 LSLAN-----LQVLEHFOQBEKAMYSAELEKQKQIAEWKKAENLEKGVISLQECLE 1724
Qy 1531 EYEKKSPKIAK--GWTNIHPDQKEYPNPQKLPENLYNLVLNQPWKVYLYNSSD----- 1582
Db 1725 EANAALDSASRLTQLOVKEEQIBELQRNELRQEMLDV--QKMLMSLANSSEGVKDV 1782
Qy 1583 FITNLFV-----EPGSDR-----GSGTKLKQVIQKQVNNYADWGA--YLTFWY---- 1626
Db 1783 LMRNLFGHFHTPKNQREHVLRLMGSILGVRRREMEQLPHD--DOGSVTRMTGWLGGGS 1840
Qy 1627 ----DKNIITNPVITANTADVFIKDVKELENDTKLIAPNIT 1665
Db 1841 KSVPTPLRPNQGVNNSSELSFVKFL-ETESHPSIPPPKLS 1882

RESULT 30

US-10-335-977-8087
; Sequence 8087, Application US/10335977
; Publication No. US2004005279A1
; GENERAL INFORMATION:
; APPLICANT: DOUGLAS SMITH et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES
; RELATING TO HELICOBACTER PYLORI FOR
; DIAGNOSTICS AND THERAPEUTICS
; NUMBER OF SEQUENCES: 10031
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LAHIVE & COCKFIELD
; STREET: 28 State Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02109-1875
; COMPUTER READABLE FORM:
; MEDIUM TYPE: CD/ROM ISO9660
; COMPUTER: IBM PC Compatible
; OPERATING SYSTEM: Windows NT 4.0
; SOFTWARE: UNIX
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/335,977
; FILING DATE: 30-Dec-2002
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/993,002
; FILING DATE: 17-DEC-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Mandragouras, Amy E.
; REGISTRATION NUMBER: 36,207
; REFERENCE/DOCKET NUMBER: GTN-018
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)227-7400
; TELEFAX: (617)742-4214
; INFORMATION FOR SEQ ID NO: 8087:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2384 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein

; HYPOTHETICAL: YES
; ORIGINAL SOURCE:
; ORGANISM: Helicobacter pylori
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (B) LOCATION 1...2384
; SEQUENCE DESCRIPTION: SEQ ID NO: 8087:
US-10-335-977-8087
Query Match 2.9%; Score 283.5; DB 15; Length 2384;
Best Local Similarity 18.5%; Pred. No. 9e-07;
Matches 394; Conservative 298; Mismatches 638; Indels 795; Gaps 105;
Qy 41 QGVISQLGID----SVAFKPSI--ANFTSDYSQSVKALLNGKTFDPKSEFTDFVSKFD 94
Db 537 QG-LDQLGFNNNDHPKKEIFEPSLGTGKFTAHAPSDKNYRFITGTELDPIISANLSKFLVPNQ 595
Qy 95 FLTNNGRTVLEIPKPYQVVIS-----EFSBEDDKE-----RFLGFHLKLEKEDG- 139
Db 596 VIQNTALENYQFYQYDAFVGNPPYGNHKKIYSSNDKELSNESIHNHYFLGKAIKELKDDGI 655
Qy 140 -----NIAQSATKFTYLLPLDMPKA----- 159
Db 656 GAFVSSWFMDAKNPKMREHIAKNAT---FLGAILRPNVFPKATGAETSDIVFPKKGVE 712
Qy 160 -----ALGOYSYIVDKNFNNL---IIHPLSNFSAQSIKPLALTRSSDFTIAKLNQFN 207
Db 713 KATNQSFTKAMPYV---DKILNSLDDDTLFAIQNNRFDSPFIP-----SDQLKIVNAVA 762
Qy 208 N-----QDELWYVLEKFFDLEALKANIRLOTADFSF--EKGNLVDVPVYFIRNPNQKE 260
Db 763 NHFGKQEKLRWYK-----IDTANFGYSTQDYKIIKDFI-----DKV 801
Qy 261 WASDLNODQKTVRLYLRTFSPQAKTI-----LKQYKQDETFLSSI---DLKASNGTSL 312
Db 802 GKNSINLNEQTLINEY----FIHPENILGHLSLEKTRYRFEINGEIQYKVDLQALDESL 857
Qy 313 FANENDLKQDLVDLDDVSDYFGGSETITANSQVKVPASERSLKDRVKFKDQKQKPRI 372
Db 858 DLSQ-ALKQAIEKLPKDVYQY----HKTTAKTDVLIIDSSNERYQEVQKLIKNI-----L 906
Qy 373 EKFSLYEYDALSFYSQLE-----LVSKP-----NSIKDLVNAVL- 407
Db 907 ERRELKVDNL--YFQLEQNNEMGIFLKPFTAINSKVQDSRLKAYFKIKALNDLTSAELN 964
Qy 408 ----ARNLRFSLGKNFLFDLASHLDYFVLSVAKIKOSSITKKLFIELPKISLKSII 463
Db 965 PLSSDLELENKRAKLNLYVDFVKFGY-----LNENKNRK-----DIRQDL 1006
Qy 464 LGDQEPNIKTLFEKEVTFKLDNFRDVEIEKAFGLLYPGVNEELEQARKAQASFEKEKSK 523
Db 1007 YGAKVLGLEKDFEKEITPRSAKQNIQPRQA-----QAKKAQ-IPFFETLNP 1052
Qy 524 KGLKFSQOKEENSKAINNOEGLEED--DNTERLPENSPYQYQENAGLGSAPDKPYM 580
Db 1053 KKELIITNAKEALIASINQKGLDLHFIRDHFTTOSLETT----- 1092
Qy 581 IKDVQNR--YILAKSQIQELIKAKDYTKLAKLSNRHTYNIISRLKEQLFDVNPRISSR 639
Db 1093 IKELLEQKLIYKDHKONGGVIANDY-----LSG-----NVKREKKE----- 1129
Qy 640 DIEKAKFVLDTKTEKNKYWOIYSSASVPFQNKWSLFGYYRYLLGLDLPQTHLVLKLGQKA 699
Db 1130 -----VKEAIN----- 1135
Qy 700 GLQFEGEYENLPDSFNLEDLKNIRIKTPLFSQKDNPKLSLLDFNNYYDGEIKAFEGFLPLF 759
Db 1136 ----QGVLEGL--EANKVDLE----- 1151
Qy 760 LPKELRRNSSGGGSONSPWEQEIIISQFKDQNLNQDLAQFSTKIWEKIIGDE-NEF 818
Db 1152 IPKDLKATEIMA-----NINSPW---IPTQYL-----EEFLMELSANHYEKQYGDQMTDY 1198

Qy 819 DONRLQYKLDLOESWINKTRONLWYVYLDKLVKP-----KNNLE 862
Db 1199 QLSNLKEDIKIEHLSGAYEVVRNNELNELYGIRHKOPHSYKVPFESLLNKLNNKDL 1258
Qy 863 AKFQISNLQELLTAFTVSAALSNNMNYQDSGAKSTIIEEIAELDPKVKVKGADVQ 922
Db 1259 VKYAQVDPNDPKKIFITDEQSN-----LARQAEELKEAFKOWIYKD-YS 1304
Qy 923 LKPHYAIGFDNAGKFNQEVIRSSRTIYLYKTSKSKLEADTID---OLNQAVQNA----- 975
Db 1305 RRTHLEQIYNDT---FNNSV-----LXTYDGSQLELEGFNYHISLAPHOKNAIFRT 1352
Qy 976 -----PLGLOSFLYDTERFGVFOKLATSLAVQH----- 1003
Db 1353 IQDRAVCLDHQVAGKTLTATASCMEQKRMGLNVR--TLIAVPNHLTKQMGDEFYKAYPN 1410
Qy 1004 -----KQEKTKPLKLN--DGYTLHDKL-----KPIVPIQISS-----SPE 1039
Db 1411 ANVLVDSKDTTEKERELLFNOIANNVYDAVIAHLELLSNPRGIIIEELKEELVNAE 1470
Qy 1040 KDMFEGKL---NONGSQNVNVSTFGSIIIESPYFTNPOEDADLDQDQ---DDSRQGNNS 1094
Db 1471 KNFERQELAYKNPRETKPNERAFKNKLD---KIRAKYDAILEKQSGSHIDISQMGIDN 1526
Qy 1095 LDNQFAGLLK-----OKLAILLGNQ-----FI--QYQOQNDKEIEF-----EII 1131
Db 1527 LIVDEAHLFKNALAFETSMKIA-GLGNQOQSGNRARDLFKTRYLHQNDKKIMFLTGTPIA 1585
Qy 1132 NVEKYSELSPRVEFKLAKTLEBNGKTIIRVLSDETMSLIVNTTIETKTPMSAVPEVFTKW 1191
Db 1586 N--SUSEMTHLQRYLTPDVLKERGL-----EFFF--DW 1614
Qy 1192 VEQY-----DPRTPLAAKT-KFVLKFKDQI PVDGSGNISTDKWLASIPLVHQQMLRSP- 1244
Db 1615 AKTYGEVNDPELDTSAQSYKMNVRFSKFSVDVQ---LSTMYRAPADIVSNDILKGNPH 1671
Qy 1245 -----VVKTIIE-----LGLTEQOQOQOQO---QOQOQPKAKVKEELETY 1285
Db 1672 FVPKYVGDKPINVVVRSEEAQFTGVALENGKYNESIIDRMQCEKSKQKQD----- 1727
Qy 1286 NPKDEFNINPLTKAHLTLIS-NLVNNDPNYKIEDLKVIKNBAGHQLAFSLRANNIKRL 1344
Db 1728 -----NILSCTTDARKVALDYRLI--DPAKVEK-----EFSKSYAMAKNIYENY 1770
Qy 1345 MNTPTIPADYNPFYFNEDWRSIDKYLNKNGVSSHQOQAGGNOGSGLIQRLNKNIPKE 1404
Db 1771 LETHATKGTOLGFI-----GLSTPKTHSQKVS-----LEALDHAHETE 1808
Qy 1405 TTPPALIALKDRNN---TNLSNYSKIIIMIKPKYLVERSIGVYPWSTGLDGYIGSQTKDG 1461
Db 1809 NKNPL-----DKAQELLESLSYDEKGNLIAP-----SKKELEN 1842
Qy 1462 TSSSSQKGFODFTQALGLKNTKHYHGLGSLIRIFDPGNELAKIKDASNKKGEEKLJKS 1521
Db 1843 ELKEKEAKSVNDEEIAKCSFDVSDVLRHLVQMGIQNEIAFIHDA--KTEE---QK 1896
Qy 1522 YDLFKNYLNEYKK---SP-KIAGKWTNI-----HPDQ-----KEY 1553
Db 1897 QDLFKK-LNRGGVRVLGSPAKMGV-TNVQERLVAMHEDCPWRPDELLQMEGRGIRQ 1954
Qy 1554 PNPQKLPENYLNVL-----NOPW-----KVTLYNSDFITNL 1587
Db 1955 NILHQNDPENFRMKIYRYATEKTYDSRMWQIITETSKGIEQFPAHKLGLNELEDF--NM 2012
Qy 1588 FVEPEGSRRGSGTK-----LKQV-IQOVNNNYADWGSAYLTFWYDKNIITNQPNV 1637
Db 2013 -----GSSNASEKKAETAGNPLIIIEVKLRAIKSEESKYKAFKNEYFNESLKN----- 2063
Qy 1638 ITANIADVPIKOVKELED-NTKLIAPNITQ-WMPNITSGSKERY-----KPTVFFGN 1687
Db 2064 -NASKLDYLKQELKOLETLQORSVIIPTHTIEIKLYDLKNEESKDYLKIKVEPL----- 2117
Qy 1688 WENENSSMNSQAQTPWEKIRE-----GFALQALKSSFDQKTRTFVLTNAPLPLW 1738

Db 2118 --KENASMEEL---THKKLKEQNKQIAEQNKEDLAIKKQFASNLTNTLTFVNEEDYKLL 2172
Qy 1739 KYGPLGF-----QNGPNPKTDWRLLVPOND----- 1763
Db 2173 EY--KGFVNAVYKTKYQVEFSLPKDNPIAYSPSNMVMYKNDTINMFSSYNFCAEIKFDG 2230
Qy 1764 -----DNQIAALRVQEOODREPKSSSE 1783
Db 2231 FLKRLDNAITKL-----PELIKE 2248
RESULT 31
US-09-820-843A-73
; Sequence 73, Application US/09820843A
; Publication No. US20030039963A1
; GENERAL INFORMATION:
; APPLICANT: Council of Scientific and Industrial Research
; TITLE OF INVENTION: A COMPUTATIONAL METHOD FOR THE IDENTIFICATION OF CANDIDATE PROTEIN
; TITLE OF INVENTION: USEFUL AS ANTI-INFECTIVES
; FILE REFERENCE: Q63915
; CURRENT APPLICATION NUMBER: US/09/820,843A
; CURRENT FILING DATE: 2001-03-30
; NUMBER OF SEQ ID NOS: 118
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 73
; LENGTH: 1805
; TYPE: PRT
; ORGANISM: M. genitalium
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: gi|1045905
US-09-820-843A-73
Query Match 2.9%; Score 281; DB 10; Length 1805;
Best Local Similarity 19.2%; Pred. No. 8.3e-07;
Matches 361; Conservative 282; Mismatches 627; Indels 614; Gaps 93;
Qy 77 KTFDPKSS-----EFTDFVSKFDFLTNNGRVTLEIPKQYVVVISFSPEDDKERPRLG 129
Db 2 KPFDPKPSLOPIYDIGDDGVLQSEYEKNRSKT--DVDKIENQLTKEIKSLEDELKNLKG 59
Qy 130 FHLKEKLEDGNGIAQSAATKFIYLLPLDMPKAAALGOYSYIVDKNFNL-----IIHPLSNF 183
Db 60 --LNQAEADN-----PE-----LDKKINHLEVDLNLVNEYKNF 91
Qy 184 SAQSIKPLALTRSDFTAKLNFQNNQDELWYLFKFFDLEALKANIRLOTADFSEK--GN 242
Db 92 QFQ-----KNHMVDKVSLEDN-----LTFYKNELTR--LOQENADFLNSKYAN 133
Qy 243 LVD-----PFVYSFIRPNQOKEMASDLNQ-----DQKTVRLYLRTEPSQAKTILKDYK 292
Db 134 LANFOANYHNKLDNFHRLIENQNTIIRLNQKINGNQNLDNNVALLQNPNITVEKQNY- 192
Qy 293 YKDETFLLSSIDLKASNGTSLFANENDLKQDLVD--LLDVSDFYFGGQSETITNSQVKPV 350
Db 193 -----LLNVID-----QLY--NEL-DQLENQKRLLSI-EYENTYRELVSADNELQNV 235
Qy 351 PASERSLKRVKVKKQKQKPIREKPSLYEYDALSFYSQLOELVSKPNSIK-----DIV 403
Db 236 --YENIDQNIQKHQVQYRDELSQLERKIQLT---KQELVDKESALRVKIDDDADFYI 289
Qy 404 NATLARNLRFSLGKYNFLFDLASHLDYFVLSAKIKQSSITKTKLFIELPKISLKSSI 463
Db 290 NARLAB-----LDDVAKQLSP-----ODGITKQNAOHVEDKL-----VA 323
Qy 464 LGDOBPNIKTLFEKEVTFKLDNFRDVEITEKAFGLLYPGVNEEELQAKAQRASFEKSKS 523
Db 324 LNKEKDRLNT--QKEAFNLQSQALIDINK-----LOQENLFAKHLEHQNEPEKQSD 376
Qy 524 KGLK-----BFSQOKBENSKAINNQE-GLEEDDNITRLPENSPFIQYQENA 569
Db 377 SILKLETKALQHKINEFKNESATKSBELLNQBELFEKREIDTLLTOASLEYEHOR- 435


```

; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 64245
; LENGTH: 1818
; TYPE: PR1
; ORGANISM: Mycoplasma pneumoniae
US-10-282-122A-64245

```

```

Query Match          2.9%; Score 280.5; DB 15; Length 1818;
Best Local Similarity 19.6%; Pred. No. 96-07;
Matches 310; Conservative 261; Mismatches 628; Indels 383; Gaps 66;

QY 56 KPSIANFTSD-----YQSVKALLNG-KTFDPKSSFTDFV--SKFDFTLNNGT 102
DB 323 KASLANLTKEKERSAERLNTALNDINRMEQENALFAKLEQQOYEFERKQOES 382
QY 103 VLEIP---KYYQVVISSEPDDEKFRGLGFLHKELE-----DGNIAQSAKTFYLLPL 154
DB 383 LLLKLETHKQLOKRIGEFKIESEAKSEALLIOERELLEKREIDDLTQASLEYEQORRT 442
QY 155 DMPKALQGYIVDKNFNNLIHPLSNFSAQSIKPLALTRSSDFTAKLNQFNNODELWV 214
DB 443 NQ---VLKEKHQVQOHQNL-VHAKKK-----LDQKHYLAEOKRI---DEQOI 485
QY 215 Y-----LEKFF-----DLEALKANIRLOTADS-----FEKGNLVDPP 247
DB 486 FKLKEKATIRERLEKLYLVKKQKQDKQDKENDLLIFEKQLRQYQADFENEIEBKQN--ELF 543
QY 248 V-----YSFIRNPQ---NOK-----EWA-----SDLN-----ODQKT 271
DB 544 ASQKSLQKSFQTKNKEALNQAKIAEDWAHLKQNHKHHADLEIFLEGEFNHLQOEKH 603
QY 272 VRLYARTESPQAKTILDKYKDETFL-----SSIDLKASNGTSLFANENDLKQDLVDL 327
DB 604 KULEARTQDNVSLLSARFKQKQAEVLVQKQSLQLEQLTAAFNKEQEAVERDKDEL--ANL 662
QY 328 LDVSDYFGQSQSETITNSQVFPASERSLKDVRFKDQKQPKRIEKFSLFYDALSFFVS 387
DB 663 EKQKEMGLDKVHQFDENSINISKLAEREL--AIKFKEKELEAAQKQSLDNNNAGLKL 720

```

```

388 QLOELVSKPNSIKDVLNATLARNLRFSLGKYNFLFDDLASHLDYVFLVSKAKIKQSITK 447
721 QLDKLSLSEKTERLELEASKEKILDP-----YDESSRIADYSDLDQARLAEVKTL 772
448 KLFIELPIKISLSKSSILGDOEPNIKTLFEKVTFFKLDNFRD-----VEIKAFGLLYPGV 502
773 K-----NQETAAKS--ERELKVALEKLNQAKKAFQIRKQQLLEIASV 814
503 NEELEQAR---KAQASFEK--EKSXKGLKEFSSQKEENSKAINN-----EGLEEDDNIT 553
815 KQQAQKANLLKQQAELDKQTELEAAFLQDQTDKKELEKALHSHVSKOELLERERSFL 874
554 ERLPENSPIQYOENAGLGASPDKPYMIKDVQVQRY--YLAKSQIQELIK-----601
875 -----LQQRFEFAEHVAGFKQVHFKTQWQLSEFNKQQQSEQIKRETELKIAFAD 926
602 -AKDYTKLAKLLGNRHTYNTISLRKE-----QLFDVNPRIPISSRDIEKAEFVL 648
927 LKKDY-QLFELQKNQEQFQIEQKHKELELLAQKAELEQKATALASQDQDTVQAKL 985
649 DKTEKNKYWQIYSSASPVFQNKWSL-FGYRYLLGLDPKQTIHELKLGKAGL-----701
986 DLARQQHELELRQNAF-----NQASLSLNKQREQJTNQVVLHGLKGRHEKLTUKDRLLA 1041
702 -----QFEGYENLPSPDN-----LEDLKNIRIKTPLFSQKDNFKLSLDD 740
1042 EKEKQHKHDAEINQRFQFENEYADQAKRELOELNQR-----RNLEQSNASILK 1095
741 FNNYDGEIKAPEFGLPLFLPKELRRNSNBSGQNSNSPWEQEIISQFQD-QNLSNQ--797
1096 KRN-----QLTDLFALLRKVQHTQTNRVQLNTQI---KEFLLEKKNFKQASDEAA 1143
798 -----DQLAQFSTKW---EKIIGDENPFDONRLQYKLLKDLQSWINKTNDNLYWT 847
1144 LQKALLIKRLSPASKLOREALAIOKLEPDKD-----EQKSEINNAKLQ-----1192
848 YLGDKLVKPKNNLEAKFRQISNLQELLTAPYTSAAALSNWNNYQDSGAKSTIIFEETAE 907
1193 ---EQFKLEKQNFDEAKQKLIIEFKDQCORLDVEKLLKQKLVQKLNLSKSVLTYKNRAD 1249
908 LDPKVEKVGADVYQLKPHYAIGFDNAGKFNQEVIRSSRTIYLYKTSKSKLEADTDDQ 967
1250 LSQOQLQHKYANLLELK-----EKLOTAKRALDKKHRAIYGMKQAFYSE 1293
968 LNOAVKQAPLGLQFYLDTER-FGVFQKLAHSLAVQHKQKEKTLPKLNNNDGYTIHDKL 1026
1294 LRQEKQQL-LSAQKQVDDKSRLLQONQHLQNLSETKKKQSLQSLHDK-----FQOR 1346
1027 KKPVIPOISSSPKDFEGKLNQ-NGOSQNV--NVSTFGSIIESFYSTNFQEDADLQD 1083
1347 RKEAVSSILNSHK-----KLKQKEGELQGIQLKSLAKKTQIEQ-FSKLYQOREKLDRO 1399
1084 QGDSRQGNNSLDNOAGLLKQKLAILLGNQFIQYQONDKEIEFEIINVEKVSLSFRV 1143
1400 RTTSLKLHRELKAQNEATAHKNREVLTEIEN---YKKELQRLATTEKSEFNNKNRLEFY 1455
1144 EFKLAKTLEDNGKTRIVLSDETMSLIVNTIETKPEMSAVPEVDFTKVVEQVDPRTPLAA 1203
1456 FRKIRNETEKKEAHIKTVLEETQK-----KRLHVEATEAKLHLQKQSIIS 1500
1204 KTKFVLKFKDOIIPVDGSGNISDKVILASIPLVHQMLRLSPVVKTIIRBLGLKTEQOQQOQ 1263
1501 KQOELKEIKERSVDISH--TNKQREELNSLLHQNLKQKLAEREREINNKDSLTQKI 1558
1264 QQQQQQQQPKKA-----VRKSEE-----LETYN-----PKDEFNINPLTKAHL 1303
1559 QTAQKQKSEKARILKLEKRAVEQQYQASITRLKTRNADLEKNDKNDKHLFPPLPKIN--1616
1304 TSLNVLNNDPNY-----KTEDLVKIKNEAGDHQALAFSLRANNIKLMMNTPIITFA 1352
1617 -----GNDMNYFPYPMFPYQOQED-----SSNQIRHLFEQOLQFM 1653
1353 DYNPPFFYNE--DWRSIDKYLNNK-GNVSSHQQAAGGNOGSGLIQRLNKI-----KPE 1404

```

Db 1654 QOR---YENELTELRRORALLEKKLDQIOLESQLSAKNDKFEKVEQMMQKLEKTEQKLS 1710
QY 1405 TFTPALIALKDRNNTNLSYSD 1426
Db 1711 AFDOKINALAEQINTQKAHAD 1732

RESULT 35
US-10-732-923-3306
; Sequence 3306, Application US/10732923
; Publication No. US20050108791A1
; GENERAL INFORMATION:
; APPLICANT: Egerton, Michael D
; TITLE OF INVENTION: TRANSGENIC PLANTS WITH IMPROVED PHENOTYPES
; FILE REFERENCE: 38-15(52796)C
; CURRENT APPLICATION NUMBER: US/10/732,923
; CURRENT FILING DATE: 2003-12-10
; PRIOR APPLICATION NUMBER: 10/310,154
; PRIOR FILING DATE: 2002-12-04
; NUMBER OF SEQ ID NOS: 24149
; SEQ ID NO 3306
; LENGTH: 1818
; TYPE: PRT
; ORGANISM: Mycoplasma pneumoniae
US-10-732-923-3306

Query Match 2.9%; Score 280.5; DB 17; Length 1818;
Best Local Similarity 19.6%; Pred. No. 9e-07;
Matches 310; Conservative 261; Mismatches 628; Indels 383; Gaps 66;

QY 56 KPSIANFTSD-----YQVKKALLNG-KTFDPKSEFTDFV--SKPDFLTNNGRT 102
Db 323 KASIANLTKEKRLSAEKQFERLNTALNINRMEQENALFAKHLQEQOYEPERKQES 382
QY 103 VLIEP---KKYQVVISPEPDKERFLGFLHKEKLE-----DGNIAQSATFIYLLPL 154
Db 383 LLKLETHKQKRIQEFKIESEAKSEALLIQERELLEKKREIDDLTQASLEYEQORRT 442
QY 155 DMPKAALGOVSYVDKNFNLLIHPLSNFSQAQSIKPLALTRSSDFTAKLNQFNQDELWV 214
Db 443 NQ---VLKEGHRVQOHQNL-VHAKKK-----LDQERHYLAQKRI---DEEQI 485
QY 215 Y-----LEKPF-----DLKALKANIRLOTADFS---FEKGNLVDPF 247
Db 486 FKLKEIATERRELEKLYLVKKQKQDKQKENDLLIFEKQLRQYQADFENEIEEKQ--ELF 543
QY 248 V-----YSFTRNQ---NOK-----EWA-----SDLN-----ODQKT 271
Db 544 ASOKSIQKSTQLKNKEAELNQAQIAEDWAHLKQNGHHADLEIFLGEFPHLQOEKH 603
QY 272 VRLYLRTFSPQARTILKDYKDETFL-----SSIDLKASNGTSLFANENDLKQDLDVL 327
Db 604 KLEARTQFNVRVSLLSARFQQAELVKQKQSLQQLTAAFNKEQEAVERDWKRL-ANL 662
QY 328 LDVSDYFGGQSETITNSQVQVPASERSLKVKKQKQKQKPRIBKFSLYVDALSFYS 387
Db 663 EKQKEMLGDKVHQFDENSLNISKLAEREL--AIKPKKEKELEAAQKQLSLDNNNAGLKL 720
QY 388 QLOELYVKPNSIKDLVNATLARNLRSGLKYNFLFDDLASHLDYFLVSKAKIKQSSITK 447
Db 721 QLDKLSKTERLELEASKERTLDF-----YDESSRRIADYSDDLQARLAEVKTLE 772
QY 448 KLFELPIKISLSSILGDQEPNIKTILFEKVTFKLDNFRD-----VEIEKAFGLLYPGV 502
Db 773 K-----NQETAAS--ERELKVALEKLNQAKKAFQIRKQQLLEIASV 814
QY 503 NEELEQAR---KAQASFEK--EKSXKGLKFEQSQKEENSKAINN-----QEGLEEDDNIT 553
Db 815 KOOLAQKANLLKNQAELDKQTELEAAFLQEDTDKKELEKALHVSYSKQELLERSFEL 874
QY 554 ERLPENSPYOQENAGLGASDPKPYMIKDQVQRY-YLAKSQIQELIK-----601

Db 875 -----LQKQREFAHVAGFKRQVHFKTQMRQRLSEFNKQOQSEQIKRETELKTAFAAD 926
QY 602 -AKDYTKLAKLSNRHTYNISRLKE-----QLFDVNPRIIPSSRDIEKAKFVL 648
Db 927 LKDY-QLFELQKQEFQOIQKHKELELLAQKAEKQLEQKATALASQDQDTVOAKL 985
QY 649 DKTERKNYQIYSSASPVFQNKWSL-FGYRYVLLGLDPKQTTHELKVKQKAGL-----701
Db 986 DLARQOHELELRQNAF---NOASLSLNKQREQLTNQVVLHGELKKRHEKLTLDORLLA 1041
QY 702 -----QFEGVENLPDFN-----LEDLKNIRIKTFLFSQKONFKLSLLD 740
Db 1042 EKEKQHKDAEINQRFQFENEYADFOAKRELELQNR-----RNLEQSNASLLK 1095
QY 741 FNNYDGEIKAPEGLPLFLPKELRRNSSGGSQNSNSPWEQEIISQPKD-QNLSNQ--797
Db 1096 KRN-----QLTLDFAALLKRVQHTQTNRVQLNTQI---KEFLEKKNFQKASDEAA 1143
QY 798 -----DOLAQFSTKIW---EKIIGDENEPONNPLQYKLLKDLQESWINKTRDNLWT 847
Db 1144 LQKALLIKELRSFASKLQOREALATQKLEFKRD-----EQKSEINNAKLQ---1192
QY 848 YLGDKLKVKPKNLEAKFRQISNLQELLTAFYTSAAALSNNVYQDSGAKSTIPEIEAE 907
Db 1193 ---EQFLEKQNPDEAKQKQLEIEFKDQCORLDVEKRLKQKLVOLKNLSKSYLTYNRAD 1249
QY 908 LDPKVKKEVGADVYQLKPHYAIGFDDNAGFQNEVIRSSRTIYLYKTSQSKLEADTIDQ 967
Db 1250 LSOQQLQHKYANLLELK-----EKLQAKRALDKHRAIYGMQAFVSE 1293
QY 968 LNOAVQNPALGOSFVLDTER-FGVFQKATSLAVQHKQKEKTLPKLNNDDGYTLTHDKL 1026
Db 1294 LRQEKQL--LSAQKQVDDKSRLLSEQNRHLNLSSETKKKROSLHDINK-----FDOR 1346
QY 1027 KKPVPQIPISSPEKDWFEKLNQ-NGOSQNV-NVSTFGSIIESPYFTSNFOEDADLDD 1083
Db 1347 RKEAVSSILNSHK-----KLQKQEGELQIIQKLSKKTQIEQB-FSKLYQOREKLDRO 1399
QY 1084 QGDSRQGNNSLDNOBAGLLKQKLAILLGNQFIQYQQNDKEIEFIIINVEKVSLSFRV 1143
Db 1400 RTTSLKLHRELKAQNEATAHKNREVLEIEN---YVKELQRLTTEKSEFDDNNKRLFEY 1455
QY 1144 EFKLAKTLEDNKTIRVLSDETMSLIVNTTIEKTPEMSAPVEFDTKWEQDPRTPLA 1203
Db 1456 FRKIRNEIEKKAHTKVLLEETQK-----KRHLVETEAVKLHLQKQSIIS 1500
QY 1204 KTKFVLKFDQIPVDGSGNISDKWLASIPLVIHQOMLRLSPVVKTIRELGLKTEQOQOQ 1263
Db 1501 KQELKEIKERSDRDISH--TNKQREELNSLHQNKLQKNLAEREREINNKSLSLTQKI 1558
QY 1264 QOQOQOQPOKKA-----VRKEBE-----LETYN-----PKDEFNINPLTKAHLR 1303
Db 1559 QTAQKQKSEKARILKLEKRAVEQOYQAHITRLKTRNADLEKNDKHLFPPLPKIN--1616
QY 1304 TSLNVNDPNY-----KTEDLVKNEAGDHQALAFSLRANNIKRLMNTPTIFA 1352
Db 1617 -----GNDMNPYPYPPWFPYQKQED-----SSNQLRHLEFQOQLQM 1653
QY 1353 DYNPPFYNE--DWESIDKYLANK-GNVSSHQOAGGQSGQGLIORLNKNI-----KPE 1404
Db 1654 QOR---YENELTELRRORALLEKKLDQIOLESQLSAKNDKFEKVEQMMQKLEKTEQKLS 1710
QY 1405 TFTPALIALKDRNNTNLSYSD 1426
Db 1711 AFDOKINALAEQINTQKAHAD 1732

RESULT 36
US-10-335-977-8088
; Sequence 8088, Application US/10335977
; Publication No. US20040052799A1
; GENERAL INFORMATION:
; APPLICANT: DOUGLAS SMITH et al

Mon Nov 21 16:44:21 2005

TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES
RELATING TO HELICOBACTER PYLORI FOR
DIAGNOSTICS AND THERAPEUTICS

NUMBER OF SEQUENCES: 10031

CORRESPONDENCE ADDRESS:

ADDRESSEE: LAHIVE & COCKFIELD

STREET: 28 State Street

CITY: Boston

STATE: Massachusetts

COUNTRY: USA

ZIP: 02109-1875

COMPUTER READABLE FORM:

MEDIUM TYPE: CD-ROM ISO9660

COMPUTER: IBM PC Compatible

OPERATING SYSTEM: Windows NT 4.0

SOFTWARE: UNIX

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/10/335,977

FILING DATE: 30-Dec-2002

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/993,002

FILING DATE: 17-DEC-1997

ATTORNEY/AGENT INFORMATION:

NAME: Mandragouras, Amy E.

REGISTRATION NUMBER: 36,207

REFERENCE/DOCKET NUMBER: GTN-018

TELECOMMUNICATION INFORMATION:

TELEPHONE: (617)227-7400

TELEFAX: (617)742-4214

INFORMATION FOR SEQ ID NO: 8088:

SEQUENCE CHARACTERISTICS:

LENGTH: 2440 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

HYPOTHETICAL: YES

ORGANISM SOURCE:

ORGANISM: Helicobacter pylori

FEATURE:

NAME/KEY: misc feature

LOCATION: (B) LOCATION 1...2440

SEQUENCE DESCRIPTION: SEQ ID NO: 8088:

US-10-335-977-8088

Query Match 2.9%; Score 280.5; DB 15; Length 2440;

Best Local Similarity 18.5%; Pred. No. 1.4e-06;

Matches 395; Conservative 316; Mismatch 678; Indels 697; Gaps 100;

QY 41 QGVISQLGID---SVAPKPSI--ANFTSDYQSVKALLNGKTFDPKSSSEFTDFVSKFD 94
DB 593 QG-LDQLGFNDNHPKKIPEPSLGTGKFTAHAPSDKNYRFICTELDPISANLSKFLYPNQ 651
QY 95 FLTNGRTVLEIPKKYQVVIS-----BFSPEDDKE-----RFLGFLHLEKLEDG- 139
DB 652 VIQNTALENYQFYQETDAFVGPNPYGNHKKIYNSDNKELSNESIHNYFLGKAIKELKDDGI 711
QY 140 -----NIAQSAATKFIYLLPLDMPKA----- 159
DB 712 GAPVVSSWFMDAKNPKMREHIAKNAT---FLGAIRLPNSVFKATGAETVSDIVFFKKGVE 768
QY 160 -----ALGOYSYIVDKFNFL---IIHPLSNFSAOSIKPLALTRSSDDFIKLNQFN 207
DB 769 KATNQSFTKAMPYY-----DKILNSLDDDTLFLALQNNRFDSPF-----SDQLKIVNAVA 818
QY 208 N-----ODELWVLEKPFDEALKANIRLOTADFSF--EKGNLVDPFVYSFIRNPONQKE 260
DB 819 NHPGFQKELQRYEK-----IDTANFGYSTQDYKIIKDFI-----DKV 857
QY 261 WASDLNQDQKTVRLYLRTFSPQAKTI-----LKQYKDETFLLSI-----DLKASNGTSL 312
DB 858 GKNSINLNEQTLENY-----FIHPENILGHLSLEKTRYRFETNGEQIYKYDQLALEDESL 913
QY 313 FANENDLKOQLDVLDDVSDYFGQSQETITSSQVFPVPSASRLKDRVFKFKDQKQKPRI 372

Db 1777 KSQKQD-----NLSCTTDARKVALDVRLL--DPNAKVEK-----BFSKS 1815
Qy 1334 FSLRANNIKRLMTPITFADYNPFYYNEDWRSDTKYLNKNVNSVSHQQAAGNQGSL 1393
Db 1816 YAMAKNIYENYLETHATKQGLGFI-----GLSTPKTHSQKVS----- 1853
Qy 1394 IQRLNKNIKBETTPALIALKDRNN---TNLSVSKIMIKPKYLVERSIGVPMWSTGLD 1450
Db 1854 LEALDNAHETENKNPL-----DKAQELLESLSYDEKGNLIAP----- 1891
Qy 1451 GYIGSEQTKDGTSSSQKGFDDFOALGLKNTYHGKLGSLIRIFDPGNEIAKIKDAS 1510
Db 1892 -----SKLENELEKEAKSVNLEDEIAKGCSDVYSVLRHLVQMGPONEIAFIHDA- 1946
Qy 1511 NKKEEELKLSYDLFNKYNLEYEKK---SP-KIAKGWTNI-----HPDQ- 1550
Db 1947 --KTEE--QKQDLFKK-LNRGGVRVLGSPAKGVG-TNVQERLVAMHELDQWPRDEL 1999
Qy 1551 -----KEYPNQKLPENYLNVL-----NQPW-----KVT 1576
Db 2000 LQMEGRGIQGNILHQNDPENFRMKIYRYATEKTYDSRMWQIIETSKSGIEQFRNAHKL 2059
Qy 1577 LYNSSDPITNLVFEPEGSGRSGTK-----LKQV-IQKOVNNYADWGSAYLTFWY 1626
Db 2060 LNELEDF--NW-----GSSNASEKABATGNPLIIEVKURAEIKSEESYKAFNKEHYF 2112
Qy 1627 DKNIITNQPNVITANIADVPIKOVKELED--NTKLIAPNITQ-WMPNISGSKFXY----- 1679
Db 2113 NEESLKN-----NASKLDYLKQELKOLETLQRSVLIPTHTTEIKLYDLKNEESKDYELIKV 2167
Qy 1680 ---KPTVFFGNWENENSSNQATPTWEKIRE-----GFALOALKSSFDQKTRTF 1727
Db 2168 KEVEPL-----KENASMSEEL---THKKLKEQNKOIARQNKELDAIKKQFASNLNTL 2217
Qy 1728 VLTTNAPLPLWKYGLGFG-----QNGPNFKTDWRLVFNOD----- 1763
Db 2218 FVNEEDYKLEY--KGFVNNAKYKQVFEFSLSPKQNPNIAYSPSNWYKNDTINWFSS 2275
Qy 1764 -----DNOIAALRVQDRPEKSSE 1783
Db 2276 YNFCABIKPFGFLKRLDNAITKL-----PEKIKE 2304

RESULT 37

US-10-732-923-3353
; Sequence 3353, Application US/10732923
; Publication No. US20050108791A1
; GENERAL INFORMATION:
; APPLICANT: Edgerton, Michael D
; TITLE OF INVENTION: TRANSGENIC PLANTS WITH IMPROVED PHENOTYPES
; FILE REFERENCE: 38-15(52796)C
; CURRENT APPLICATION NUMBER: US/10732,923
; CURRENT FILING DATE: 2003-12-10
; PRIOR APPLICATION NUMBER: 10/310,154
; PRIOR FILING DATE: 2002-12-04
; NUMBER OF SEQ ID NOS: 24149
; SEQ ID NO 3353
; LENGTH: 1478
; TYPE: PRT
; ORGANISM: Plasmodium yoelii yoelii
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(1478)
; OTHER INFORMATION: unsure at all Xaa locations
US-10-732-923-3353

Query Match 2.98; Score 279.5; DB 17; Length 1478;

Best Local Similarity 18.3%; Pred. No. 7.6e-07; Matches 313; Conservative 268; Mismatches 607; Indels 519; Gaps 76;

Qy 353 SERSLKDRVFKKDOQKPRKEKFSLEYDALSFSYLOELVSKPNSIKDLVATLARNLR 412
Db 21 SRKSRESXVRENADD-----ENLAIKERELINQLNELNIISEKEMHSDAI-----KNMK 70

Qy 413 FSLGKYNFLDLDLASHLDYFVLVSKAKIKQDSITKKLFIELPIKISLSSILGQBPNI- 471
Db 71 HQLDEFQ-----DQNNYI 84
Qy 472 --KTLPEKEVTPKLDNFRDVEIEKAGFLYPGVNBELEQAKAQRASPEKESKKGJKEF 529
Db 85 NEURLIEKLDLTKRD-----EIEK-----FNKLEBINRREESLAYKE-----NEF 125
Qy 530 SQOKEENKAINNOEG--LEEDDNITERLPENSPIQ---YQENAGLIGASPKDPYMIKDV 584
Db 126 IQIKEE-----INNRERDIIDREHEITEKYNELSYKEKELVEKEHELKNESDILKNEDF 181
Qy 585 QNORYYLAKS-----QIQEL-----IKADYTKL-----AKLLSN 614
Db 182 INNNEELARKREELNIEQIQIBELKNDLILKFEHLHNEQDIKIKE-TKLKTYIKKLLKN 240
Qy 615 RHTYNISLRLKEOLF-----DVNPRIPSSROI-----EKAKFVLDKTE 652
Db 241 NINDLLDVSDENLFSRTRSNINGNENENSNISMSMKIAELDNLENQSENINLDSLE 300
Qy 653 KKKYWIY---SSASPVQNKWSLFGYYRYLLGLDP-----KOTIHLVLKL 695
Db 301 YTYTFNEFKKIKNLKQDLQSRKELEFFETKLEIEKREKEEYRNELDNKLLHAEQLN 360
Qy 696 GQKAGLQFGEYNLPSPDFNLEDLKNIRIKTLPFSQKDNFKLSLLDPNNYVDGEIKAPEFG 755
Db 361 KKKLDDQLBRYKN--DDEHI--IKSLKESEEIINEKNTLILEL-----QQKLAQASYE 409
Qy 756 LPLFLPEKLRRSSNGSGSONSPWEQEIISQFKDNLNSQDQLAQFSTKIWEKILGDE 815
Db 410 ISMIENKSNKNNNSYNGER-NSEYEKKL-----EELNN-----ITNYSKEINEL 455
Qy 816 NBFQDNRLQY--KLLKDLQESWINKTRDNLWYTLGDK---LKVPKKN-----L 861
Db 456 NKEKENITQPFMEKIMDD--EEGINKLEE---LNDANALIVLNKNKNELNLYNVIEM 508
Qy 862 EAKFROISNLQELLTAFYTSAAALSNWNNYVQSGAKSTIIFEBIA----- 906
Db 509 EQANKDMRDDIDILLA--NIDKLNDEKNTVEKEQELKYNELKINIEHKYKCEKNFFN 566
Qy 907 ELDPKVKVKGADVOLKPHYAIGFDDNAGKNOEVISSRTIYLKTSKSKLEADTID 966
Db 567 MLPPKMKKI---EYKGRNSKEGNNINMSQNSSVISKDHESL-----GNSSEETCDID 619
Qy 967 QLNQAVKNAPLGLQSPY---LDTERFGVFKLATS LAVQHKKQKTLPKKLANDGYTLI 1022
Db 620 KIVNVVNGQNGKRASIKNINILNEKDMVIBSLNEKLLYVEKDQAKN--HEIINDYKSKI 677
Qy 1023 HDKLLKPVIPQISSSPEKDWFEKG-----LNQOSQNV-----NVSTFGSIE 1066
Db 678 DELLSKLN-----FEGNMPFFIKINEIENNENYTHQYVLNLLNAILYRKIIT 726
Qy 1067 SPYFSTNFOEDADLDQGDGDSRQGNNSLDNQA---GLLKOKLAILLGNQPT----- 1116
Db 727 KIFIENQFYKEIIQIKYEIFDIISKNTSKKKTETKLSMIKKGSPKGGHSHSNYPL 786
Qy 1117 -----QYYQNDKEIEFEIINVEKVSSELSPRVEFKLAKTLEDNGKTIIRVLSDETMSL- 1168
Db 787 HMSDAIHKYQNPXDEQSGMDGLIEPTNW-----IIKAVNDEPDTFENICNELYNLN 839
Qy 1169 -----IIVNTTIEKTPMSAVPEVFDTKWVQYDPRTPLAOK 1204
Db 840 KSEEMDIMYKLVNLEYSYNSIINIINEIKTKNISNI-----EKSELKKN 886
Qy 1205 TKFVLKFKDOIIPVDGSGNISDKWLASIPLVHQQMLRL---SPVVKTI-----RE 1251
Db 887 MKLLKKKYNLSNDFQONVEDVDKLLILVKESEQNELLVTENEELKNLYKELNDEYNEK 946
Qy 1252 LGLKTEQQOQQOQQOQQOQQOQKAVRKEBELETYPNKBENFLNPLTKAHLRLTSLNVNN 1311
Db 947 LNLIKQNEYQIKNLQBLIEKENENKTKTEINEBF-LKTDLDYLTLSLEQANQSLTNLSE 1005

Qy	1312	DPNTKI-----EDLVKIKNEAGDHOAFSLRANNIKELMNTPTITPADYNNPFFYYN-EDW	1364
Db	1006	NEKNKIALKOLTEENIYLKNOIEDKEBENIEYLTOKIKSNDQVISELKEFEMLIKVKETY	1065
Qy	1365	RSIDKYLNNKGNVSHQOAGAGNGGSLI--ORLNKNKIKPETETPALIALKADRNNNLS	1422
Db	1066	DSIELNKRKGN--NHTKGHIENNSTTLYDHNFYEDIDKVELLGWISKLENDNN---1119	
Qy	1423	NYSKIMIKPKYLVERSIGVPMSTGLDGYTGE-----QTKDGTSSSSQOQKFDQ	1473
Db	1120	NLKEECMLKN-----DFYILSEKNHELEBIIQKNDLPIKYNDELDIKN	1163
Qy	1474	DFIQALGLKNTYHCKLGLSIRIPDPGNELAKIKDASNKKEGLKLSYD-----1523	
Db	1164	DIIEKENLINERNKYKI-KCIQIID-----ICFNKDFSIIDIREKIVAFENDDEEINI	1218
Qy	1524	-----LFXNYLNE-----YEKKS-----PKIAGWTNI--HPDQKEYPNPNKLPENYL	1565
Db	1219	NCHRSMLNNTNEEFGYKKNPNKPNARHSRIIRGTSKVCVADNNQY-NENESMKORLS	1277
Qy	1566	N-----LVL--NOPWKVTLX-----NSSDIFITNLV	1589
Db	1278	NORKDDAAKTHOENITHEIEMSKONEELIKKNEQIKWLKNIYIEDLNKEITNKDYII-MKS	1336
Qy	1590	EPGSDRSGTKLKOVIOKQVNNVADNGSAYLTFWYDK-NIITNOPNVITANTADYFIK	1648
Db	1337	QOESADK-----NDVLEKNQYIIFLKDOIKLKCN--NIDENNLFDI---1376	
Qy	1649	DVKELEDNKTLPAPNITQWPNPISGSKFKYPTVFFGNGWENENSSMNSQAQPTWEKIR	1708
Db	1377	-----DNVNL-SPTLR-----SFIKKSYNNRCSFVEVDEKSNHGHE-----N	1415
Qy	1709	EGFA-LQALKSSFDQ--KTRTFVLTTNAPLPLMKYGLPLGFGNGPNFKTQDWRLVFQNDN	1765
Db	1416	ENFASRQRRRTFFDOSRKRLSEFTS-----QEOYNF-----YDEN	1451
Qy	1766	QIAALRVOEQDRPEKSSDKOKKMIK	1792
Db	1452	EV---KIEBMNLKMDLNDKLKQKRNK	1475
RESULT 38			
US-10-661-809-23			
; Sequence 23, Application US/10661809			
; Publication No. US2004010191A1			
; GENERAL INFORMATION:			
; APPLICANT: HOOK, Magnus			
; TITLE OF INVENTION: BIOINFORMATIC METHOD FOR IDENTIFYING SURFACE-ANCHORED PROTEINS FR			
; FILE OF INVENTION: POSITIVE BACTERIA			
; FILE REFERENCE: P07741US01/BAS			
; CURRENT APPLICATION NUMBER: US/10/661,809			
; CURRENT FILING DATE: 2003-09-15			
; PRIOR APPLICATION NUMBER: 60/410303			
; PRIOR FILING DATE: 2002-09-13			
; NUMBER OF SEQ ID NOS: 24			
; SOFTWARE: PatentIn version 3.1			
; SEQ ID NO 23			
; LENGTH: 10203			
; TYPE: PRT			
; ORGANISM: Staphylococcus epidermidis			
US-10-661-809-23			
Query Match 2.8%; Score 275.5; DB 16; Length 10203;			
Best Local Similarity 18.5%; Pred. No. 2.1e-05;			
Matches 375; Conservative 319; Mismatches 758; Indels 573; Gaps 93;			
Qy	74	LNKGT-----FDPKSEFTDFVSKFDELTNNGRTVLETPKKYQVVISFSPEDDKERPR	127
Db	3189	LNMGTOESINNYTTKEREAGNIASSADTIINNGDASIE-----QITEN-----3231	
Qy	128	LGPHLKEKLEDGNIASATKFIYLLPLDMPKALGOYSYIVDKN-----FNNLIITH	178
Db	3232	-----KIRVEEATNALNEAKQHLTADTTSLKTEVRKLRSGDTNNKPSVSAAYNN-TTH	3285

Qy	179	PLSNFSAQSIKPLALTFSSDFTAKLNQFNQNDDELWVYLEKFFDLEALKANIRLOTADFSF	238
Db	3286	SLOSEITQT-ENRANTIINKPIRSVEVNNALH-----EVNQLNQLT-----3327	
Qy	239	EKGNLDVPFYVSPIRPNQKQEWASDL-NQOQKTVLYLRTFSPQAKTILKDYKYKDET	297
Db	3328	DTINLQP-----LANKESLKEARNLESKINETVQTDGTMTOQS-----VENYK-----3371	
Qy	298	FLSSIDLKASNGTSL---PANENDLKDQ-----LDVDLLDVSDYFGQSESTITSNQ	346
Db	3372	---QAKIKAGNESSIAQTLINNGDASDQEVSTIEKLNQKLSBETN-----SINHLTYNKE	3424
Qy	347	VKVPASERSLKORVKPKDQKQPKRIEKFSLYSDALSFYSQQLQELVSKPNSKNDLVNAT	406
Db	3425	--PLETAKNOLOANI-----DQKPSD--CMTQOSVQSYERKLQEAQDKINSI---NNV	3471
Qy	407	LARNRFLSGKYNFLFDDLASHLDYVPLVSKAKIKOSSITTKKFIPIELPIKLSKSSILGD	466
Db	3472	LANNPDVNAIRTN-----KVETEQIINNEL-----TOAQKGLTVD	3505
Qy	467	QEP--NIKTLEK-----EVTFFLDNPRDVEIEKAFGLLYPGVNEELEQAR-K	511
Db	3506	KQPLINAKTALQOSLDNQPSITGTGTEATIQYNNAKROKAEQ-----VIONANKIITENAOPS	3561
Qy	512	AORASPEKSKKGLKEFSQ-----QKEENSKAINNOEGLEDDDNITERLPENSPI--	562
Db	3562	VQVSDSEKSVQEQALSSELNNAKSALRADKQELQOAYNQ---LIQPTDLNNKPPASITAYN	3618
Qy	563	-QYQENAGLGAS-----PDKPYMIKDVQNOYRYLAKSQIOELIKAKDYTKLAKLLS	613
Db	3619	QRYQOFSNELNSTKTNTDRILKEQNPVSADVNN-----ALNKVREV--QOKLNEARALLQ	3671
Qy	614	NRHTYNI SLRLKEQLFDVNPRIPSS-----RDIEKAKFVLDKTEKN	654
Db	3672	NKEDNSALVRAKEQLQAVDQVPSPTGEGTQOTKDDYNSKQAAQOESKAAQVINDGAT	3731
Qy	655	KYMQIYSSASPVFQNKWSLFGYRYLLGLDIPKOTIHELKVLGQKAGLQFEGYENLPSDFN	714
Db	3732	T--QOISNAKTNVERALEALNNAKTGLRADKELOAYNQLTQ-----NIDTSGKTPASIR	3785
Qy	715	LEDLKNIRIKTPLFSQKDNFKLSLLDFNNYD-----GEIKA--PEGLPLFLPKELRR	766
Db	3786	KYNEAKSRIQTQIDSAKNEANSILTNDPQVSQVTAALNKIKAVQPELDKAIAMLNKEN	3845
Qy	767	NSSNSGGSQ-----NSNSPWE--QELISFKQONLSNQDLAQFSTKIWEKII--GDE	815
Db	3846	NNALVQAKQQL-QQIVNEVDPTQGTMTDTANNYKSKKREADEIQKA-----QOILNNGDA	3900
Qy	816	NE---FDQNNRLOYLKLLKOLQESMINKTRDMLYWTYLGDKLKVPK-----NNLEAKPRQ	867
Db	3901	TEQQITNETNRVNOAI-----NAINKAKNDL-----RADKSQLEWAYNQLTQVDTNGKK	3950
Qy	868	ISNLQELLTAFYTSAAALSNMNNYQDSKAKETIIFEEIAELDPKYKEKVG-----DVYQL	923
Db	3951	PASIQOYQAA---RQAITQYN-----NAKSE-AHQIILENSNPSVNEVAQALQKVEAYOL	4001
Qy	924	KFHYAIGFDDNAGKPNQEVIRSSRTIYLTGSGKSKLEADTIDQLN-----QAVK	973
Db	4002	KVNDIAIHILQN--KENNSALVTAKNLOQSVNDQPLTTGTMQDSINNYEAKRNEAQAIR	4059
Qy	974	NAPLGLQ-----SPYLDTERFVGFKLATSFLAVQHKKQKEKTLPKKLNNDGYTLIHDKLKK	1028
Db	4060	NAAEVINNGDATAKQISDEKSKVQALA-----HLNDAKQOQLTADTTTEL	4103
Qy	1029	PVPIQISSSPKDWPEGLKNGQSONV---NVSTFGSIIIESPYFSTNFQEDADLDQDQ	1085
Db	4104	QTAVO-----QANRGDTNNKPPRSINAYNKAIOSL-----4134	
Qy	1086	DDSRGNNSLNQEAGLLKQKLAILLGNQFTQYQONDKEIEBEIINVEKYSE-----L	1139
Db	4135	--ETQITSKDNANAVIQPIRTVQEVNNALQQVNNALNQLNQLTEAINQLPSLNNDAKAA	4192

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 28, 2005, 20:50:32 ; Search time 217 Seconds

(without alignments)
3348.953 Million cell updates/sec

Title: US-10-607-631-8

Perfect score: 9732
Sequence: 1 MNKNKSTLLATAAAIIGST.....TNNAFNNVKEFNISKIVE 1879

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 2105692 seqs, 386760381 residues

Total number of hits satisfying chosen parameters: 2105692

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_16Dec04.*
1: Geneseqp1980s.*
2: Geneseqp1990s.*
3: Geneseqp2000s.*
4: Geneseqp2001s.*
5: Geneseqp2002s.*
6: Geneseqp2003as.*
7: Geneseqp2003bs.*
8: Geneseqp2004s.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	9598	98.6	1879	8	ADI28615 Mycoplasma
2	9435.5	97.0	1878	8	ADI28627 Mycoplasma
3	475	4.9	1032	8	ADI28611 Mycoplasma
4	415	4.3	1001	8	ADI28617 Mycoplasma
5	325	3.3	4688	6	ABU48941 Protein e
6	305.5	3.1	1875	6	ABR53560 Protein s
7	305.5	3.1	1875	7	ADK64380 Disease t
8	305.5	3.1	1875	8	ADs43855 Bacterial
9	303	3.1	2633	4	ABG06505 Novel hum
10	302	3.1	2663	4	AAM39097 Human pol
11	302	3.1	2688	4	AAM40883 Human pol
12	301	3.1	2663	4	ADQ17932 Human sof
13	299	3.1	2017	4	ABG06301 Novel hum
14	295	3.0	1639	2	Aaw54145 P. falcip
15	295	3.0	1639	5	Aae29345 Plasmodiu
16	295	3.0	1654	1	AAP50777 Sequence
17	291.5	3.0	896	2	AAR05591 The 36 kD
18	291	3.0	5005	6	ABU48947 Protein e
19	290.5	3.0	1679	8	ADs43650 Bacterial
20	290.5	3.0	1979	3	AAB18171 Plasmodiu
21	288	3.0	1948	5	ABP73774 Candida a
22	286.5	2.9	1979	7	ADU75595 Prostata
23	281	2.9	1805	6	ABU35589 Protein e
24	281	2.9	1805	7	ABO23572 Mycoplasma
25	281	2.9	2441	3	AAB18161 Plasmodiu

26	280.5	2.9	1818	6	ABU36321 Protein e
27	280.5	2.9	2440	2	Aaw20828 H. pylori
28	276	2.8	3259	7	Ades6037 Human pro
29	276	2.8	3259	7	Ades6033 Human pro
30	275.5	2.8	1663	2	AAR46608 Plasmodiu
31	275.5	2.8	10182	5	ABP38314 Staphyloc
32	275.5	2.8	10182	8	ADO84851 S. epiderm
33	275.5	2.8	10203	8	ADO84803 Staphyloc
34	274.5	2.8	1997	3	AAB18287 Plasmodiu
35	273.5	2.8	1690	4	ABB61144 Drosophil
36	273.5	2.8	1690	4	ABB61173 Drosophil
37	272	2.8	1516	3	AAB18195 Plasmodiu
38	271	2.8	3225	7	ADJ68448 Human hea
39	270	2.8	6641	6	ABU42656 Protein e
40	269.5	2.8	1464	8	ADP25450 Plasmodiu
41	262.5	2.7	2295	3	AAB18180 Plasmodiu
42	262.5	2.7	2681	6	ABU19025 Pathogen
43	262.5	2.7	5024	4	AAG82935 S. epider
44	262	2.7	3418	7	ADJ68372 Human hea
45	261.5	2.7	5171	7	ADJ70881 Human hea

ALIGNMENTS

RESULT 1

ADI28615
ID ADI28615 standard; protein; 1879 AA.

XX AC ADI28615;

XX XX 22-APR-2004 (first entry)

XX DE Mycoplasma hyopneumoniae immunogen protein C28-MHP545.

XX KW Pneumonia; vaccine; diagnosis; C28-MHP545; immunogen; immunostimulant;
XX KW antiinflammatory.

XX OS Mycoplasma hyopneumoniae.

XX XX WO2004003161-A2.

XX XX 08-JAN-2004.

XX XX 27-JUN-2003; 2003WO-US020460.

XX XX 28-JUN-2002; 2002US-0392632P.

XX XX (IOWA) UNIV IOWA STATE RES FOUND INC.
XX XX (NSWA-) NSW AGRIC.

XX PI Minion CF, Mahairas GG, Djordjevic SP;

XX DR WPI; 2004-083044/08.

XX DR N-PSDB; ADI28614.

XX PT New immunogenic Mycoplasma hyopneumoniae polypeptide, useful in eliciting
XX PT an immune response and in treating or preventing enzootic pneumonia.

XX PS Claim 5; SEQ ID NO 8; 81pp; English.

XX CC The present sequence is the protein sequence of C28-MHP545, an
XX CC immunogenic polypeptide from Mycoplasma hyopneumoniae strain 232. The
XX CC invention provides M. hyopneumoniae polypeptides and nucleic acids
XX CC ADI28608-ADI28627, including C28-MHP545, as well as vectors and host
XX CC cells. Compositions containing M. hyopneumoniae polypeptides and nucleic
XX CC acids are provided for use in methods of treating pigs to prevent
XX CC enzootic pneumonia. Diagnostic tests for detecting M. hyopneumoniae
XX CC infection in swine herds are also provided.

XX SQ Sequence 1879 AA;

Query Match 98.6%; Score 9598; DB 8; Length 1879;

Best Local Similarity 98.8%; Pred. No. 0; Matches 1857; Conservative 8; Mismatches 14; Indels 0; Gaps 0;			
Qy	1	MKNKSTLLATAAAIIGSTVFGTVGLASKVKYRGVNPQTQGVISQGLGIDSVAFPSTA	60
Db	1	MKNKSTLLATAAAIIGSTVFGTVGLASKVKYRGVNPQTQGVISQGLGIDSVAFPSTA	60
Qy	61	NFTSDYQSVKALLNGKTFDPKSESEFTDFVSKDFLTNNGRVTLEIPKKYQVVISFSP	120
Db	61	NFTSDYQSVKALLNGKTFDPKSESEFTDFVSKDFLTNNGRVTLEIPKKYQVVISFSP	120
Qy	121	DKERRFLGHLKEKLEDGNIASATKFIYLLPLDMPKALGOYSYIVDKNFNNLIHPL	180
Db	121	DKERRFLGHLKEKLEDGNIASATKFIYLLPLDHPKALGOYSYIVDKNFNNLIHPL	180
Qy	181	SNFSAGSIKPLALTRSSDFIAKLNQFNQDELAVYLEKFPDLKALNARLQTAQSF	240
Db	181	SNFSAGSIKPLALTRSSDFIAKLNQFNQDELAVYLEKFPDLKALNARLQTAQSF	240
Qy	241	GNLVDPVYFIRNPQNKESWADLNQDQKTVRLYLRTPEFSPQAKTILKDYKYKDETFLS	300
Db	241	GNLVDPVYFIRNPQNKESWADLNQDQKTVRLYLRTPEFSPQAKTILKDYKYKDETFLS	300
Qy	301	SIDLKASNGTSLFANENDLKDQDLVDLLDVSDFEGGQSEITITSNSQVKVPASERSLK	360
Db	301	SIDLKASNGTSLFANENDLKDQDLVDLLDVSDFEGGQSEITITSNSQVKVPASERSLK	360
Qy	361	VKPKDQOKPRIEKFSLYEYDALSFYSQLEYSKPNISKDLVNATLARNRSLGKYNF	420
Db	361	VKPKDQOKPRIEKFSLYEYDALSFYSQLEYSKPNISKDLVNATLARNRSLGKYNF	420
Qy	421	LFDDLASHLDYFLVSKAKIKQSSITKKLFIELPIKISLSSILGQDQEPNITLPEKEVT	480
Db	421	LFDDLASHLDYFLVSKAKIKQSSITKKLFIELPIRISLSSILGQDQEPNITLPEKEVT	480
Qy	481	PKLDNFRDVEIEKAFGLLYPGVNEELEQARAKARASFEKSKKGLKFSQOKEENSKAI	540
Db	481	FRLDNFRDVEIEKAFGLLYPGVNEELEQARAKARASFEKSKRGLKFSQOKEENSKAI	540
Qy	541	NNQEGLEEDDNIITERLPENSPIQYQOENAGLGLASPKPYMIKDVQNRVYLAKSQLELI	600
Db	541	NNQEGLEEDDNIITERLPENSPIQYQOENAGLGLASPKPYMIKDVQNRVYLAKSQLELI	600
Qy	601	KAXDYTKLAKLNRHTYNTISLRKQGLFDVNPRIIPSSRDIEKAKFVLDKTERKNWQIY	660
Db	601	KAXDYTKLAKLNRHTYNTISLRKQGLFDVNPRIIPSSRDIEKAKFVLDKTERKNWQIY	660
Qy	661	SSASPVQNKWSLFGYYRYLLGLDPKQTHLVLKQKAGLOFEGYENLPDNLKLN	720
Db	661	SSASPVQNKWSLFGYYRYLLGLDPKQTHLVLKQKAGLOFEGYENLPDNLKLN	720
Qy	721	IRIKTPLFSQKDFKLSLLDFNNYDGEIKAPEFGLPLFLPKELRRNSSNGSGQNSNP	780
Db	721	IRIKTPLFSQKDFKLSLLDFNNYDGEIKAPEFGLPLFLPKELRRNSSNGSGQNSNP	780
Qy	781	WEQEIISQFKDQNLNQDQLAQSTKIWEKIGDENEFDONNRLOYLKLDQESWINKT	840
Db	781	WEQEIISQFKDQNLNQDQLAQSTKIWEKIGDENEFDONNRLOYLKLDQESWINKT	840
Qy	841	RDNLWYTYLGDKLKVKPKNNLEAKFRQISNLQELLTAFYTSALSNNNWNYQDSGAKSTI	900
Db	841	RDNLWYTYLGDKLKVKPKNNLEAKFRQISNLQELLTAFYTSALSNNNWNYQDSGAKSTI	900
Qy	901	IFBEIAELDPKVKKEKVADYVQLKFHYAIGFDDNAGKFNQEVIRSSSRITYLKTSGSKL	960
Db	901	IFBEIAELDPKVKKEKVADYVQLKFHYAIGFDDNAGKFNQEVIRSSSRITYLKTSGSKL	960
Qy	961	EADTIDOLNOAVKNAPLGLQSFYLDTERFGVFOKLATSLAVQHKQEKTLPKKLNDGYT	1020
Db	961	EADTIDOLNOAVKNAPLGLQSFYLDTERFGVFOKLATSLAVQHKQEKTLPKKLNDGYT	1020
Qy	1021	LHDKLKKPVIPOISSPEKDFEGKLNQNGSQSNVNVSTFGSIIESPVFSTNFQDADL	1080
Db	1021	LHDKLKKPVIPOISSPEKDFEGKLNQNGSQSNVNVSTFGSIIESPVFSTNFQDADL	1080
Qy	1081	DQGGQDSDRQGNNSLDNQEAGLLKQKLAILLGNQFIQYIQQNDKEIEFBIINVEKVSLS	1140
Db	1081	DQGGQDSDRQGNNSLDNQEAGLLKQKLAILLGNQFIQYIQQNDKEIEFBIINVEKVSLS	1140
Qy	1141	FRVEFKLAATLBDNGKTIIRVLSDENSLIIVNTTIEKTPEMSAPVEFDFTKVEQYDPRTP	1200
Db	1141	FRVEFKLAATLBDNGKTIIRVLSDENSLIIVNTTIEKTPEMSAPVEFDFTKVEQYDPRTP	1200
Qy	1201	LAAKTKFVLKFDQIIPVDGSGNISDKWLASIPLAVIHQOQMLRSLPVVKTIRELGLKTEQQ	1260
Db	1201	LAAKTKFVLKFDQIIPVDGSGNISDKWLASIPLAVIHQOQMLRSLPVVKTIRELGLKTEQQ	1260
Qy	1261	QQQQQQQQQPKKAVRKEELETYNPKDEFNLNPLTKAHLTLTSLNVLNNDPNYKIEDL	1320
Db	1261	QQQQQQQQQPKKAVRKEELETYNPKDEFNLNPLTKAHLTLTSLNVLNNDPNYKIEDL	1320
Qy	1321	KVUKHAGDHQAFSLRANNIKRLANTPITFADYNPFFYYNEDWRSIDKYLNNKGNVSSH	1380
Db	1321	KVUKHAGDHQAFSLRANNIKRLANTPITFADYNPFFYYNEDWRSIDKYLNNKGNVSSH	1380
Qy	1381	QQAAGGNOGSGLIQRLNKNIKPETFTPALIALKDRNNTNLSYSDKIIMIKPKYLVERS	1440
Db	1381	QQAAGGNOGSGLIQRLNKNIKPETFTPALIALKDRNNTNLSYSDKIIMIKPKYLVERS	1440
Qy	1441	IGVPWSTGLDGYIGSEQTKDGTSSSSQQKGFQDQFIQALGLKNTYHGKLGSLIRIFDPC	1500
Db	1441	IGVPWSTGLDGYIGSEQTKDGTSSSSQQKGFQDQFIQALGLKNTYHGKLGSLIRIFDPC	1500
Qy	1501	NELAKIKDASNKKGBEKLLKSYDLFKNYLNEYEKSPKIAKGTWNIHPDQKEYPNPQKL	1560
Db	1501	NELAKIKDASNKKGBEKLLKSYDLFKNYLNEYEKSPKIAKGTWNIHPDQKEYPNPQKL	1560
Qy	1561	PENYLNVLNQPKWTLVNSDFITNLFVEPEGSDRGSGTKLKQVTKQVNNNYADWGA	1620
Db	1561	PENYLNVLNQPKWTLVNSDFITNLFVEPEGSDRGSGTKLKQVTKQVNNNYADWGA	1620
Qy	1621	YLTFWYDKNIIITNPQNVITANIADVFIDKVKLEDNKTLIAPNITQWPNISGSKEKPYK	1680
Db	1621	YLTFWYDKNIIITNPQNVITANIADVFIDKVKLEDNKTLIAPNITQWPNISGSKEKPYK	1680
Qy	1681	PTVFGNWNENENSNMQATPTWEKIREGPAQALAKSSFDQKTRFTVLTNAPLWKY	1740
Db	1681	PTVFGNWNENENSNMQATPTWEKIREGPAQALAKSSFDQKTRFTVLTNAPLWKY	1740
Qy	1741	GPLGFQNGFNFTQDWRLVFQNDNQIAALRVQEOORPEKSSDKQKWKIKFVVIPEE	1800
Db	1741	GPLGFQNGFNFTQDWRLVFQNDNQIAALRVQEOORPEKSSDKQKWKIKFVVIPEE	1800
Qy	1801	MFNSGNIRFVGWYQIGPNLMLPVINSSVYDFYRGTDGSDNDVANLNPVAPQVKTIAFT	1860
Db	1801	MFNSGNIRFVGWYQIGPNLMLPVINSSVYDFYRGTDGSDNDVANLNPVAPQVKTIAFT	1860
Qy	1861	NNAFNNVFKFNISKKIVE 1879	
Db	1861	NNAFNNVFKFNISKKIVE 1879	
RESULT 2			
ADI28627			
ID	ADI28627	standard; protein; 1878 AA.	
XX	ADI28627;		
AC	XX		
DT	XX		
DT	22-APR-2004	(first entry)	
XX	XX		
DE	XX	Mycoplasma hyopneumoniae immunogen protein C28-MHP545.	
DE	XX	Pneumonia; vaccine; diagnosis; C28-MHP545; immunogen; immunostimulant;	
KW	XX	antiflammatory.	
KW	XX		
XX	XX		
OS	XX	Mycoplasma hyopneumoniae.	

XX	PV	W02004003161-A2.
XX	PD	08-JAN-2004.
XX	PF	27-JUN-2003; 2003WO-USO20460.
XX	PR	28-JUN-2002; 2002US-0392632P.
XX	PA	(IOWA) UNIV IOWA STATE RES FOUND INC.
XX	PA	(NSWA-) NSW AGRIC.
XX	PI	Minion CF, Mahairas GG, Djordjevic SP;
XX	WI	WPI; 2004-083044/08.
DR	N-	PSDB; ADI28626.
XX	PT	New immunogenic Mycoplasma hyopneumoniae polypeptide, useful in eliciting
XX	PT	an immune response and in treating or preventing enzootic pneumonia.
XX	PS	Claim 11; SEQ ID NO 20; 81pp; English.
XX	SS	The present sequence is the protein sequence of C28-MPH545, an
CC	CC	immunogenic p102 paralog from Mycoplasma hyopneumoniae strain J. The
CC	CC	invention provides M. hyopneumoniae polypeptides and nucleic acids
CC	CC	ADI28608-ADI28627, including C28-MPH545, as well as vectors and host
CC	CC	cells. Compositions containing M. hyopneumoniae polypeptides and nucleic
CC	CC	acids are provided for use in methods of treating pigs to prevent
CC	CC	enzootic pneumonia. Diagnostic tests for detecting M. hyopneumoniae
CC	CC	infection in swine herds are also provided.
XX	SS	Sequence 1878 AA;
XX	SS	Query Match 97.0%; Score 9435.5; DB 8; Length 1878;
XX	SS	Best Local Similarity 97.0%; Pred. No. 0;
XX	SS	Matches 1827; Conservative 17; Mismatches 29; Indels 11; Gaps 3
Qy	Db	1 MNKKSTLLATAAAIIIGSTVFGTVVGGLASKVKYRGVNPTQGVISQLGLIDSVAFKPSIA 60
Db		
Qy	Db	1 MNKKSTLLATAAAIIIGSTVFGTVVGGLASKVKYRGVNPTQGVISQLGLIDSVAFKPSIA 60
Db		
Qy	Db	61 NFTSDYQSVKKALLNGKTFDPKSSEFTDFVSXKFDFLTNNGRTVLIPPKYQVVISFSPSE 120
Db		
Qy	Db	61 NFTSDYQSVKKALLNGKTFDPKSSEFTDFVSXKFDFLTNNGRTVLIPPKYQVVISFSPSE 120
Db		
Qy	Db	121 DDKERFRLGFLHKEKLEDCGNIAQSATKIYYLPLDMPKAALGOYSYIVDNKFNNLIHHPL 180
Db		
Qy	Db	121 DDKERFRLGFLHKEKLEDCGNIAQSATKIYYLPLDMPKAALGOYSYIVDNKFNNLIHHPL 180
Db		
Qy	Db	191 SNFSAQSIKPLALTBSDDFIKLNQFNODELWVLEKFFDLLEALKANIRLOTADFSPEK 240
Db		
Qy	Db	191 SNFSAQSIKPLALTBSDDFIKLNQFNODELWVLEKFFDLLEALKANIRLOTADFSPEK 240
Db		
Qy	Db	241 GNLDVPFVVSYFTRNPQNKEWASDLNQDKTVRLYLRTETFSPOAKTIKDQKYKDETFLS 300
Db		
Qy	Db	241 GNLDVPFVVSYFTRNPQNKEWASDLNQDKTVRLYLRTETFSPOAKTIKDQKYKDETFLS 300
Db		
Qy	Db	301 SIDLKASGTSLFANENDLKOQLDVLLDSVDFGGQSETITTSNQQVPPVASERSLKDR 360
Db		
Qy	Db	301 SIDLKASGTSLFANENDLKOQLDVLLDSVDFGGQSETITTSNQQVPPVASERSLKDR 360
Db		
Qy	Db	361 VKFKKQQOKPRIEKFSLSLEYDALSFYSQLOELVSKPNSIKDLVNATLARNLFSLGKYNF 420
Db		
Qy	Db	361 VKFKKQQOKPRIEKFSLSLEYDALSFYSQLOELVSKPNSIKDLVNATLARNLFSLGKYNF 420
Db		
Qy	Db	421 LFDDDLASHLDYYFLVSKAKIKOSSITKKLFIELPIKISLSSILGDQEPNIKTLEKEVT 480
Db		
Qy	Db	421 LFDDDLASHLDYYFLVSKAKIKOSSITKKLFIELPIKISLSSILGDQEPNIKTLEKEVT 480
Db		
Qy	Db	481 FKLDNFRDVEIEKAFGLLYPGVNEELEQARQARASFEEKSKKGLKEFSQOKEENSKAI 540
Db		
Qy	Db	481 FKLDNFRDVEIEKAFGLLYPGVNEELEQARQARASFEEKSKKGLKEFSQOKEENSKAI 540
Db		

Qy 1052 OSQNVNVTGSIIESPYSTNFQEDADLDQGDSDRGNSLNDQEGALLKQKLAILL 1111
 Db 976 NSLNTDKLTF---LISFLNKKDKNPKDLKADNKNDSPINPT-----IARQKLIKII 1026

RESULT 4
 ID ADI28617 standard; protein; 1001 AA.
 XX AC ADI28617;
 XX 22-APR-2004 (first entry)
 XX Mycoplasma hyopneumoniae immunogen protein C28-MHP662.
 XX Pneumonia; vaccine; diagnosis; C28-MHP662; immunogen; immunostimulant;
 XX antiinflammatory.
 XX Mycoplasma hyopneumoniae.
 XX Key Location/Qualifiers
 XX Misc-difference 1 /note= "Encoded by TTG"
 XX WO2004003161-A2.
 XX 08-JAN-2004.
 XX 27-JUN-2003; 2003WO-US020460.
 XX 28-JUN-2002; 2002US-0392632P.
 XX (IOWA) UNIV IOWA STATE RES FOUND INC.
 XX (NSWA-) NSW AGRIC.
 XX Minion CF, Mahairas GG, Djordjevic SP;
 XX WPI; 2004-083044/08.
 XX N-PSDB; ADI28616.
 XX New immunogenic Mycoplasma hyopneumoniae polypeptide, useful in eliciting
 XX an immune response and in treating or preventing enzootic pneumonia.
 XX Claim 6; SEQ ID NO 10; 81pp; English.
 XX The present sequence is the protein sequence of C28-MPH662, an
 XX immunogenic polypeptide from Mycoplasma hyopneumoniae strain 232. The
 XX invention provides M. hyopneumoniae polypeptides and nucleic acids
 XX ADI28608-ADI28627, including C28-MPH662, as well as vectors and host
 XX cells. Compositions containing M. hyopneumoniae polypeptides and nucleic
 XX acids are provided for use in methods of treating pigs to prevent
 XX enzootic pneumonia. Diagnostic tests for detecting M. hyopneumoniae
 XX infection in swine herds are also provided.

Query Match 4.3%; Score 415; DB 8; Length 1001;
 Best Local Similarity 22.4%; Pred. No. 2.2e-15;
 Matches 255; Conservative 161; Mismatches 397; Indels 326; Gaps 54;

Qy 2 KKKKSTLLATAAAIIGSTVFQTV---VGLASKVKYRGVNTQGVISQLGLHDSVAKP 57
 Db 24 KKKKSTNLGRKLLTGAAVFFGIAITITPLVTVANWKIKOPRLQVQNAKLITNIQLKD 83
 Qy 58 STANFTSDYQSVKKAALLNGKTFDPKSSBPTDFVSKFDLTNNGRVTLEIP-----KKY 110
 Db 84 EYQNGNLSTFDKKQLFNADNTKKTGIDISQF---FDFYQKN-NTSLPINFATDYGWRY 139
 Qy 111 QVVVSEFSPEDDKERFLRGLHLEKLELDGNIAQS---ATKFTYLLPLDMPKAAALQOYSY 167
 Db 140 KLDVDFDLQDQEQSFEIYVRLVYQLPDDKKAISDLITQKIWNV---LPDYSLANFANF 196

Qy 168 VDKNFNNLIHPLSNFSAQSIKPLA-LTRSSDFIAKLQNFNNOELWVYLEKFFDLEALK 226
 Db 197 SSSKLEKRAYTNKEISLSTKKELTKLVKLEDFEKQVNMANNNEARKIINKYFNLEEEII 256
 Qy 227 ANIRLQTAQSPF--EKGNIWDPFVYSFIRNPQKQEWASD--LNQD-----QKTV-RLY 275
 Db 257 AEI-LNNKEFSYLDSEG-----IWNQYQIELVRDQILGQDFLAKTGOKGIYKLT 305
 Qy 276 LRTERSPQ-AKTLKDYKYKDETFLLSSIDLKASNGTSLPANENDLKQDLDDVLDSDYF 334
 Db 306 FYAASFENFAKTAADLN-KSKFHFGINILNN---LP-----LD----- 342
 Qy 335 GQOSETITNSQVPPASERSLKDRVKFKQOQKPRIEKFSYEDALSFYQOLQELVS 394
 Db 343 ----KTVAE--IKITEFSEDDYVPOINPEKLEA-----EINGWDFLNNY--NQIFA 388
 Qy 395 KPSIIDLNVNATLARNLRFSLGKYNFLFODLASHLDYFVFLVSKAKIKQSSITKLFIEPL 454
 Db 389 TQNEREDFLKNLIAKIVRTPLK-KVEFENKLSGIDY-----AKFLK-----YUKLD 434
 Qy 455 IKISLSSILGDQEPNI-KTLFEKVTFKLDNFRDVEIEKAGFLLYPGVNEELEQARQAQ 513
 Db 435 IKLDANSTKLAFNQKQIVAKIFGKIILRNAEN--QIVAEKNF-----SQTIEHLNRLG 485
 Qy 514 RASFEKEKSKGLK-BFSQOKBENSKAINNQEGLEDDNITERLPENSPIQOQENAGLG 572
 Db 486 QNDAELVKQIKQTKFEF---KPETRKKIANQK----- 515
 Qy 573 ASPDKPMIKDVQNRYYLAKSQIOELIKAKDYTKLAKLSNRHTYNI--RLKSOLF 629
 Db 516 -----APKSEILALLNANKDFLKNILENGDYYGYEFNEERLK--LL 555
 Qy 630 DVNPRIPSSRDIEKAEFVLDKTEK-----NKYV-----QIVSSASPVFQNKWSLFGVRY 679
 Db 556 VHSQLPNVVEEFKLSVVPKMSGEGIINLWKFQKTNQEVSTFSLAKRDISFVAKYVY 615
 Qy 680 -LLG-----LDPKQTIHELVLKQAGLQFEGYENLPSDFNLEDLKNIRIKTFLPSQKDNF 734
 Db 616 DLLNKKFLIDPKT-----QWPNLDQNSLFPKLSQIKIOPP---EKKAV 656
 Qy 735 KLS-----LLDFNNYDGEIKAPEGLPLFLPKELRRSSNGSGSONSPWEQEIISQPK 790
 Db 657 SLTSDFWLFSLNNDY-----LISPDYLN-- 684
 Qy 791 DONLSNQDQAQFSTKIWEKIIGDNEFDQNNRLOYKLLKDOESWINKTRDNLWYTVLG 850
 Db 685 HSNLKNLIDL-----IKTESAFNRDFVEH--IRELAKS----- 716
 Qy 851 DKLVKPKNNLEAKFRQ-ISNLOELLTAFYTSAAALSNNWNYQDSGAKSTIIFBEIAB-L 908
 Db 717 ----IKPKDFIQEKGKNPITNLSEFLAVFVELI-----YSKQDQ---LLAESLGQNL 761
 Qy 909 DPKVYK---EKVGADV-----VOLKPHYAIGFDDNAGKFNQEVIRSSS 947
 Db 762 DYKIQFELEPISLNAVVSQEKTPNPNLRLNLRKLYWYKIGSDVQNGNLQIYIOTPK 821
 Qy 948 RTIYLKTSKSKLEADTIDOLNQAQVKNAPLGQSLQFYLDTDFGVFQKLSLAVQHKQ-- 1005
 Db 822 ETDLVNNNNKLJSEDEVEKLEINATPSPADQIIFLKEDY-----TQLVDSIAKQVI 874
 Qy 1006 KEKTLPLKLNNDGYTLIHDKLKPKVPIQISBSPKDMFEGKLNQNG-----OSQNVNVS 1059
 Db 875 KTEMTFVKIDN-----QIKNLPSPQFPENNYPDYGVFIITKTSKNLESS 917

RESULT 5
 ID ABU48941
 XX ABU48941 standard; protein; 4688 AA.
 XX AC ABU48941;
 XX 19-JUN-2003 (first entry)
 XX XX

DE	Protein encoded by Prokaryotic essential gene #34468.	
XX	Antisense; prokaryotic essential gene; cell proliferation; drug design.	
KW	Ureaplasma urealyticum.	
OS	W0200277183-A2.	
PN	03-OCT-2002.	
XX	21-MAR-2002; 2002WO-US009107.	
PF	21-MAR-2001; 2001US-00815242.	
XX	06-SEP-2001; 2001US-00948993.	
PR	25-OCT-2001; 2001US-0342923P.	
PR	08-FEB-2002; 2002US-00072851.	
XX	06-MAR-2002; 2002US-0362699P.	
PR	(ELIT-) ELITRA PHARM INC.	
PA	Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW,	
XX	Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;	
PI	WPI; 2003-029926/02.	
DR	N-PSDB; ACAS2811.	
XX	New antisense nucleic acids, useful for identifying proteins or screening	
CC	PT for homologous nucleic acids required for cellular proliferation to	
CC	PT isolate candidate molecules for rational drug discovery programs.	
XX	Claim 25; SEQ ID NO 76865; 1766pp; English.	
XX	The invention relates to an isolated nucleic acid comprising any one of	
CC	CC the 6213 antisense sequences given in the specification where expression	
CC	CC of the nucleic acid inhibits proliferation of a cell. Also included are:	
CC	CC (1) a vector comprising a promoter operably linked to the nucleic acid	
CC	CC encoding a polypeptide whose expression is inhibited by the antisense	
CC	CC nucleic acid; (2) a host cell containing the vector; (3) an isolated	
CC	CC polypeptide or its fragment whose expression is inhibited by the	
CC	CC antisense nucleic acid; (4) an antibody capable of specifically binding	
CC	CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular	
CC	CC proliferation or the activity of a gene in an operon required for	
CC	CC proliferation; (7) identifying a compound that influences the activity of	
CC	CC the gene product or that has an activity against a biological pathway	
CC	CC required for proliferation, or that inhibits cellular proliferation; (8)	
CC	CC identifying a gene required for cellular proliferation or the biological	
CC	CC pathway in which a proliferation-required gene or its gene product lies	
CC	CC or a gene on which the test compound that inhibits proliferation of an	
CC	CC organism acts; (9) manufacturing an antibiotic; (10) profiling a	
CC	CC compound's activity; (11) a culture comprising strains in which the gene	
CC	CC product is overexpressed or underexpressed; (12) determining the extent	
CC	CC to which each of the strains is present in a culture or collection of	
CC	CC strains; or (13) identifying the target of a compound that inhibits the	
CC	CC proliferation of an organism. The antisense nucleic acids are useful for	
CC	CC identifying proteins or screening for homologous nucleic acids required	
CC	CC for cellular proliferation to isolate candidate molecules for rational	
CC	CC drug discovery programs, or for screening homologous nucleic acids	
CC	CC required for proliferation in cells other than S. aureus, S. typhimurium,	
CC	CC K. pneumoniae or P. aeruginosa. The present sequence is encoded by one of	
CC	CC the target prokaryotic essential genes. Note: The sequence data for this	
CC	CC patent did not form part of the printed specification, but was obtained	
CC	CC in electronic format directly from WIPO at	
CC	CC ftp.wipo.int/pub/published_pct_sequences	
XX	Sequence 4688 AA;	
SQ	Query Match 3.3%; Score 325; DB 6; Length 4688;	
	Best Local Similarity 19.4%; Pred. No. 4.5e-09;	
	Matches 450; Conservative 309; Mismatches 771; Indels 786; Gaps 118;	
QY	40 TOGVISGLGIDSVAPKPS-IANFTSDYQSVKALLNGKTFDPKSEFTDFVSKPDLTN 98	
	785 TKTVASIGLDGKAIFKTSDDAIFADPH-----KYTLTKTEADNKKVANIDEISPLDRVN 840	

QY	99	---NGRTVLEIPKQYQVWISFSPED----
DB	841	KQKNGN--VNADNKEFKIPDQKNKDLTAVYKDKGNNEIHVPIKTDDKGKVIWPNNNLF 898
QY	151	LPLDMPK-----AALGOY--SYIVDKQFN-----LIHPLSNFS 184
DB	899	-----DPNKIYDFDKIVDLNEYPNKTIIDRNSINKDVSAINDGVDQARKVLKAPAVSNVT 954
QY	185	AOSI-----KPLATRSSDFIAKLNQFNODELWVLEKF----- 219
DB	955	VNAINFQVNLFDLNLKLSYNRQFALT-----IAKVDLNDTKQ---YIAYDPTNNYKLN 1006
QY	220	FDLEALKANIRLQADTFSEKGNLVDPPFVYSFIRNPQONKEWASDINQDKTVRL----Y 275
DB	1007	FDFTHLDDANTKYVVDVEL-----LELNKEKPIKLKDDV 1042
QY	276	LRTSPQAKTILK-----DYKYK---DETFLSSIDLKASNGTSLFANENDLKDOL-- 323
DB	1043	LNFEFTSATTTINPPIWTKFDVVVKTNNDTTITFEIDK----DNILKNDQKIYAQLAL 1098
QY	324	DVDLLDVSDFG-----GQSETITSNSOVKPPVPSERSLKDVRKFKDQKQKPRIKFSL 377
DB	1099	MDDLLD--TDVINPLVFNKTNKIASINGLDLKGNSKYSIKNLYLNDQNKVYLFKNDV 1157
QY	378	YEYDALSFYSQELVSKPNSIKDLVNAATLARNLPSLCKYNFLFDDDLASHLDYFLVSK 437
DB	1158	TKYE-----QHFTNPHKI-----NLSFN--KSAVEQDIFADHANLF----- 1192
QY	438	AKIKOSSITKKLFIELPIKI-----SLXSSILGDOEPNIKIL 474
DB	1193	--IDVKDYDQKLRIKINEDVKIYYQINDTKNHLQIGYKVVANNKIKFNLVGLKE---KTT 1247
QY	475	FEKEVTFKLDNPRDVEIEKAFGLYPGVNHELEQAKAQRASFESKSKKGLKEFSQOKE 534
DB	1248	YVKKLEALNKASSIVSEFLLDPTN-----FSTSNKNTTLVGLNSIDNWN 1297
QY	535	ENSKAINNOEGLEED--DNITERLPENSPIQOQENAGLSPDKPYMKIQVONQRYILA 592
DB	1298	DQTPINAKINIGDDFQDNQVKL-----IYVSDNKEI-----KSKAVTLIKQRYN-- 1345
QY	593	KSQIQELIKAKDYTKLAKLL--SNRHTYNIQLRKEQLFDVNP-----RIPSRDIEKAK 645
DB	1346	QEFESNLIKRLYT-FSKIVYETNNQTLHLKDLTLHQ-FSINPSNNNAVSLKONTNIEITK 1403
QY	646	FVL-----DKTEKNKYWQIYSSASVPFQNKW 671
DB	1404	RLVNNDSQLISAKIEVDIDNVLNTADKPNIVYQLENSDANNLKLATNAQVVVENNK 1463
QY	672	SLFGYRYRLGLDLPQOTIHELKVLGOKAGLDFEGYENLPSD--FNLKLNRIKTKPLFS 729
DB	1464	KELKFDLVSLKINQNVYKEI-----SFNSKPTNAYFNFTNNKTNNI---VVS 1508
QY	730	QKDNKFLSI-----LDENNYDGEIKAPFEGPLPLPKELRRN 767
DB	1509	YDEQNKISLNNINPTSYKPTKNNKDSVNDVLDQVLDQLLANQY-URLKL-KQJNDN 1566
QY	768	SS-----NSGGSQ-----SNSPWEQBIISQDKQNLNSQ--DOLAOFSTKIWEK 810
DB	1567	KTVWTDPIILFNNNAKISFKLSNLHNRAVELEGLYFDDQNSVNDNTNNQISFNSKIHP 1626
QY	811	ITGDEN-----BFDQNNRLQYKLLKDLQESWINKTRDNLWYTVLGDKLKVKPKQNL---- 861
DB	1627	KIEFEPSTLTINYDTNNAIKTVSAHNAQVHFKLKTNDLEALENDQIIEAVFAPTNNDQK 1686
QY	862	--EAKFRQISN-----LQELLTAFTVSAALSNNNNYODSGAKSTIIEEITA 906
DB	1687	VVEAKLNNTNENFNEGELEFNLSGUREETTYRLIKVTFKKNKAYELLNKNGVIFE--- 1743
QY	907	ELDPVKKEVGADVYQLKPHYAIGFDDNAGKFNQEVIRSSRTIYLKTSCKSKLEADTID 966
DB	1744	-----YKNGSOAVEF-----TTQKFEHKVIDVVSSTSTNTTQOEITVK---ID 1783

Db 483 -QOYL-----LITSVNSDSKGLR-----KEEIQINIMQEDDSTITESDSQKVVT 530
Qy 217 EKFPDLKALNIRLQADPSFEK--GNLVDPPFYFIRPNQKQEWASDLNODQKTVRL 274
Db 531 ERLVEFKN-----IQOEKNAELLKVVNLDADKLESKEKSKQSIOKTESVTNKEAII 587
Qy 275 YLRTEFSPQAKTILKQYKDE--TFUSSIDLKASNGT--SLFANENDLKQDLVDLLDV 330
Db 588 TLAKSEKMDLSRIEELQELKELKTSVPNEDASYNVTIKQLTETKRDLESQVQDLQTRI 647
Qy 331 SDYFGGSEFII--SNSQVQVPASERSLKDRVFEKQDQKPRIEKSLFEDYDALSFYSOL 389
Db 648 SQITRESTENMULNKIEQIDLYSKSDISIKLGEKSSRI LAERFKL----- 695
Qy 390 QELVSKPNSIKDLVNATLARNLAFSLGKYNFLFDLASHLDYFLVSKAKIKOSSITKIL 449
Db 696 -----LSNLTDLTKAEN-----DQLRKFDY-----LQNTILKQDSKTHET 731
Qy 450 FIELPIKISLSSILGQEPNITLFEKVTFFKLDNPROVEIEK-----AFG 496
Db 732 LNEY-VSCSKSLSVETELLNLKE--EQKLRVHLEKNLQOELNKLSPKDSLRIMVTQLQ 788
Qy 497 LLYPGVNEELEQARKA--ORASFEEKSKGLKFEFSQOKEENSKAINNQBLEEDDNI 555
Db 789 TLQKEREDLLETRKSKCKIDLELDALELSEKTSQKHIIKOL-----BEDN----- 837
Qy 556 LPENSPITQOQENAGLSPDKPMIKDVQNVQRYLAKSQ--IQELIKADYTKLAKLLS 613
Db 838 ---NSNIEWQ--NKIEALKDYESVITSVDKQTDIEKQYKVYSLEKEIEEDKI----- 888
Qy 614 NRHTYNTSLBLKQOLFVNPRIPSSRDIEKAKFVLDK--TEKNKYQWQYSSASFPQNKW 671
Db 889 RLHTYNY--MDETINDDSL--KELEKSKINLTDAVSQIKYKDYKDYVETTSQSLOQT 941
Qy 672 SLFGYRYRLGLDP-----KOTIHELKVGQAGLQFEGVENLPSPFNLEDLKNIR 722
Db 942 S-----KLDESFKDFNQIKNLTDKTSLEDKISLKEQMFNLNLEDLIQ----- 986
Qy 723 IKTPLFSQKDNFK--LSLLDFNNYDGEIKAPFGLPLFLPKELRLNNSNGSGSQNSNP 780
Db 987 -KKGMEKADFKKRISILQNNKNEVEAVSEVESKLSKIQLNDLQDTIYANTAQN--- 1042
Qy 781 WEQEI-----ISQFKDQ-----NLSNQDLAQSPKIWEKIIGDENFQ 820
Db 1043 YEOELQKHADVSTISLREQLHTYKGVKQVTLNLS-RDQL-----EN----- 1083
Qy 821 NNRLQYKLLADQESMINKTRDNLWYTLGDKLVKPKNNLEAKFRQISNLQELL---TA 877
Db 1084 -----ALKENKSSWQKESLL-----EQDLS-----NSRIEDLSQNKLLYDQIQ 1125
Qy 878 FYTSAALSNNWNYQDSGAKSTIIF-----EEIAELDPKV----- 912
Db 1126 IYTAADKEVN--NSTNGPGLNILLITLRRERDILDTKVYVAERDAKMLRQKISLMDVELQD 1184
Qy 913 -KEKVGADVQLKHFVAI--GFDNAGKFNQ--EVIRSSRTI--YLTGSKSKLEADT-I 965
Db 1185 ARTKLDNSRYEKENHSHSIIQQHDDIMEKLNQLNLLRESNITLRENNNNKKELQSEL 1244
Qy 966 DQNLQAVKNAPLGL-----QSPYLDTERFGVFPQKATSLAVCHQKKEKTLPK 1012
Db 1245 DKLKQNV--APIESELTKYSMQEKQELKLAKEEVHRWKKSQDILEKHEQLSSSDYE 1302
Qy 1013 KLANNDGYTLHD-----KLKKEPVIQISSPEKDWFEGLKNQNGSQNV 1056
Db 1303 KLSSEIENKEELKENKQGAEEKFNLRQAQERLKS-----KLSQDSITEQV 1354
Qy 1057 N-VSTFGSIIIESYFTNFOEDADLDQDQDDBRQGNNSLD----- 1096
Db 1355 NSLRDAKNVLENSLEAN-----ARIEELQNAKVAQGNQNLQEAIRKLQEDAERLOAK 1410
Qy 1097 -----NOEAGLLK-----QKLAILLGNQ----- 1114

Db 1411 LBESTTSYESTINGLNEEBITTLKBEIEKORQIQOQLQATSANEQNDLSNIVESNMKKSFE 1470
Qy 1115 -FIQYYQNDKEIEFEI-----INVEKV-----SELSFRVEFKLAKTLBNDNGK 1156
Db 1471 DKIKFPIKETEQVNEKILAEQERLNQPSNINMEEIKKKWSEHEQVSKIREAEEALKK 1530
Qy 1157 TIRVLSDEMTSLIVNTTIE-----KTPEMSAVPEVFTKVVQVQDPTPLA 1202
Db 1531 RIRLPTEEKINKIIEERKEELEKEFEKVEERIKSMEQSGEIDVVLRLQLEAKVQBQKE 1590
Qy 1203 AKTKFVLKFKDQI--PVDGSGNIS-----DKWLASIPLVIIHQQLRLSPVVKTI 1257
Db 1591 LENEYNKQLEELKQVPHSSHISDDERDKLRAEISRLREEFNNELOAIKK-----KSF 1644
Qy 1258 QOQQOQQOQQOQQOQKAVRKEEEL-ETYN-----PKDEFNINLPITKAHRLTSLN 1311
Db 1645 DESGQAMKMTLLERKLAKMESQLSETKQSAESPSPKSVNNVQNPLGLPRKIEEN--SN 1702
Qy 1312 DP-NYKIEDLKVIKNEAGHQALAFSLRANNIKRLMNTPTTFADYNPFFYFNEDWRSIDKY 1370
Db 1703 SPFNPLLSGEKLLKLSKSSSGGFN-----PFTSPSPNKHQLQNDNDKR---ES 1747
Qy 1371 LANKGNVSSHQQAAGNQGSGLI---QRLNKNIKPETFTPALIALKDRNNTNLSNYSK 1427
Db 1748 LANKDPPHLEPSPNIPASRGLISSSTLSTDTNDEELTSNNPAQKDSNRNVQSEED- 1806
Qy 1428 IIMIKPKYLVERSIGVPSWSTGLDGYIGSQTKDGTSSSSQKGFQDQDFIQALGLKNTEYH 1487
Db 1807 -----TEKKKEGEP-----VKRGEAIBEQTKSNKRPIDEVGELKNDE-- 1843
Qy 1488 GKLGLSIRIFDPGNELAKIKDASNKKGBEKLKSVYDLFKNYLYNEYKKSFKPI 1539
Db 1844 -----DDTTENINESKTIKT-----EEKETDKV 1869
RESULT 7
ADR64380
ID ADK64380 standard; protein; 1875 AA.
XX
AC ADK64380;
XX
DT 06-MAY-2004 (first entry)
XX
DE Disease treating protein complex-derived protein #1185.
XX
KW protein complex; drug target; diagnosis.
XX
OS Unidentified.
XX
PN EP1338608-A2.
XX
PD 27-AUG-2003.
XX
PF 20-DEC-2002; 2002EP-00102902.
XX
PR 20-DEC-2001; 2001EP-00130253.
XX
PA (CELL-) CELLZOME AG.
XX
PI Bauer A, Gavin A, Superti-Furga G, Kuester B, Schultz J;
PI Marloeh M, Grandi P, Krause R, Kruse U, Merino A, Bauch A;
PI Michon A, Leutwein C, Rick J;
XX
DR WPI; 2003-638460/61.
DR N-PSDB; ADK64381.
XX
PT New proteins and protein complexes from eukaryotes, useful as targets in
PT drug screening, or in diagnosing or screening for the presence of a
PT disease or disorder, or a predisposition for developing a disease or
PT disorder in a subject.
PS Disclosure; SEQ ID NO 2369; 13pp; English.
XX

CC The invention relates to novel protein complexes comprising a first and a
CC second protein, or its derivative, fragment, homologue or variant. The
CC proteins are selected from given protein complexes, which are not defined
CC in the specification. The variants are encoded by nucleic acids that
CC hybridize to the nucleic acids encoding the proteins under low stringency
CC conditions. The protein complexes are useful as targets for an active
CC agent of a pharmaceutical. These protein complexes are particularly
CC useful as drugs targets for the treatment or preventing of a disease or
CC disorder. The complexes and methods above are useful in diagnosing or
CC screening for the presence of a disease or disorder or a predisposition
CC for developing a disease or disorder in a subject. These are also useful
CC in screening for a drug for treatment or prevention of a disease or
CC disorder. The molecule that modulates the amount, activity or protein
CC components of the complex is useful for the manufacture of a medicament
CC for the treatment or prevention of a disease or disorder. This sequence
CC corresponds to a protein of the invention. (Note: the sequence data for
CC this patent did not form part of the printed specification but was
CC obtained from the EPO in electronic format).

XX Sequence 1875 AA;

SQ

Query Match		3.1%; Score 305.5; DB 7; Length 1875;
Best Local Similarity		19.6%; Pred. No. 1.6e-08;
Matches 340; Conservative. 271; Mismatches 627; Indels 494; Gaps 75;		
Qy	60	ANFTSDYQSVKALLNGK-----TF-----DPKSEFTDFVSKDFL-----TNGR 101
Db	380	AKSSDFFLKKQIKERRTEHLQNIQTETVELEHKVPVINSFKERTDMLNELNAA 439
Qy	102	TVLETPKYQVVISFSDDEKPRFLGFHLEKLEDGNIASATKFIYLLPLDMPKAL 161
Db	440	LLLE-----HTSNEKNAKVELNAKQKLVCEENDLQTLK-----QLDLCR--- 482
Qy	162	GOYSYIVDKNFNNLIHPSAQSIRKPLATLRSSDFIAKLQ-----FNNDELWYL 216
Db	483	-QIQYL-----LITNSVNSDKGLR-----KBEIQFIQIMQDDSTITESDQVVT 530
Qy	217	EKFPDLEALKANIRLOTADPFEK---GNLVPFVYFTRNFQNKQEWASDLNQOKTVRL 274
Db	531	ERLVEFKNI---IQLEKNAELLKVVRLNADLKLESKEKKSKQSLQKIESETVNEAKEAI 587
Qy	275	YLRTFSPQAKTILDKYKDS---TFLASIDLKASNGT---SLFANENDLKQDLVDLDV 330
Db	588	TLKSEMDLESRIEQLQLELELKTSPVEDASYNVTIKQLETYKRLDESQVQLQTRI 647
Qy	331	SDYFGQSEITIT-SNSQVKVPVPSRSLKDRYKFKKQDQKPRIEKFSLYEYDALSFSOL 389
Db	648	SOITRESTENMSLLNKEIQDLVDSKDISIKLGKSSRILAEERFKL-----695
Qy	390	OELVSKPNSIKDVLNATLARNRFLSGKYNFLFDDLASHLDYFVFLVSKAKIKQSSITKYL 449
Db	696	-----LSNTLDLTKAEN-----DQLRKFDY---LQNTILKQDSKTHET 731
Qy	450	FIELPKISLSKSLGDQEPNITLFEKEVTFKLONFRDVELEK-----AFG 496
Db	732	LNEY-VSKSKLSIVETELNLKE---EQKLRVHLEKNLQKQELNKLSPKQSLRIMVTOLO 788
Qy	497	LLYPGVNELEOAKA-QRASPEKSKKGLKEFSQOKEENSAINNOEGLEEDDNITER 555
Db	789	TLQKEREDELETRKSCQCKIDLELDELSELKETSQKDHHLKQL-----EEDN--- 837
Qy	556	LPENSPIOQOENAGLGASDPKPYMIKOVONORYYLAQSQ---IQELIKAKOYTKLAKLIS 613
Db	838	---NSNIEWYQ-NKTEALKKQVESVITSVDSQTDIEKLQYKVKLSLEKEIEEDKI----- 888
Qy	614	NHRTYNISRLKEQLFDVNPRLPSSRDIKAKFVLDK---TEKNKYQIYSSASPYFQNKW 671
Db	889	RLHTYNV---MOETINDSLR-----KELEKSKINLTDAYSQIKEYKDLTYETTSQSLOQTN 941
Qy	672	SLFGYRYLLGLDP-----KQTIHELVLKLGQKAGLQFGYENLPSPDNLELKNIR 722
Db	942	S-----KLDESFKDFTNQIKNLTDEKTSLEDKISLKQSMFNLNELDQ-----986

Qy	723	IKTPLFSQDNFK--LSLLDFNNYDGEIKAPBFLPLFLPKELRRNSSNGSGSQNSNP 780
Db	987	-KKGMEKADFKRISILQNNNKEAVKSEVESKLSKIQNDLDQOTIYVANTAQN--- 1042
Qy	781	WEQEI-----ISQFKDQ-----NLSNODQLAQSTKIWEKIIGENEFQD 820
Db	1043	YEQLQKHADVSKTISELRQLHTYKGVQVTKLNLSDQL-----EN----- 1083
Qy	821	NNRLOYKLLKLOESWINKTRDNLWYTLGDKLVKPKNNLEAKFRQISNLQELL---TA 877
Db	1084	-----ALKENKSSWSKESL-----EQDLDS-----NSRIEDLSQNKLLYDQIQ 1125
Qy	878	FYTSAAALSNNWYQDSGAKSTIIF-----EETAEALDPKV-----912
Db	1126	IYTAADKEVN-NTNGPGLNLLITLRRERDILDTKVVAERDAKMLRQKISLMDVELQD 1184
Qy	913	-KKGADVVQLKPHVAI--GFDNAGKFNQ-EVIRSSRTI--YLTSGSKSLEADT-I 965
Db	1185	ARTKLONSRVEKENHSSIIQOHHDDIMEKLNQLNLLRESNITLNELENNNNKKELQSEL 1244
Qy	966	DOLNOAVKNAPLGL-----QSFVLDTERFGVFQKLATSLAVOHKQKEKTLPK 1012
Db	1245	DKLKQNV--APISELTALKYSMQEKEQELKLAKEEVHRWKKSQDILEKHEQLSSSDYE 1302
Qy	1013	KLNNDGVTLLHD-----KLKCPVIPQISSSPKDWPEGLKLNQNGSQNV 1056
Db	1303	KLSEIENLEKELENKERQGAEEKENRLRRQAEERLKTSS-----KLSQSLTEQV 1354
Qy	1057	N-VSTFGSIIESPFTNFQEDADLDQDQDSDSQNGNSLD-----1096
Db	1355	NSLRDAKNVLENSLSEAN---ARIEBLQNAKVAQGNNOLEAIRKLOEAEKASRELOAK 1410
Qy	1097	-----NOEAGLLK-----OKLAILLGNQ-----1114
Db	1411	LEESTTSYESTINGNEEITTLKEEIEKQRIQOOLQATSANEQNDLSNIVESMKSFEE 1470
Qy	1115	-FIQYQONDKIEIFEI-----INVEKV-----SELSFRVEFKLAKTLEDNGK 1156
Db	1471	DKIFEIKETQEVNEKLEAQERLNQPSNINWEEIKKWESEHEQVSQKIREAEALKK 1530
Qy	1157	TIRVLSDETMSLIVNTTIE-----KTPEMSAPVPEVPTDKWVQYDPTPLA 1202
Db	1531	RIRLPTEEKINKIIRKKEELEKEFEKVEBERIKSMEQSGEIDVVLRKQLEAKVQEKKE 1590
Qy	1203	AKTFVLKFKDQI-PVDGSGNIS---DKWLASIPLVHQOMLRLSPVVKTIRELGLKTE 1257
Db	1591	LENEYKKLOBELKDVPHSSHISDDERDKLRAEIESRUREEFNNELOAIKK-----KSF 1644
Qy	1258	QOQOQOQOQOQOQKQKAVRKEBEL-ETYN-----PKDEFNINPLTKAHLTLNLYNN 1311
Db	1645	DEGQOQAMKTTILLERKLAKMESQLSETKQAEPPKSVNNVQNPPLGLPKRIEEN--SN 1702
Qy	1312	DP-NYKIEDLVIRKNEAGHOLAFSLRANNIKRLMTPITFADYNPFYFNEDWRSIDKY 1370
Db	1703	SPFPNLLSGEKLKLNKSSSGGFN-----PFTSPSPNKHQLQNDNDRK---ES 1747
Qy	1371	LNNKGVSSHQQOAGAGNQGSLI---QRLANKIKPETFTTALLIALKDRNNNTNLSNYSDK 1427
Db	1748	LANKTDPPTHELPFNIPASRGLISSSSTLTDTNDEELTSNNPAQOSSNNRNOSEED- 1806
Qy	1428	IIMIKPKYLVRSIGVPWSTGLDGYIGSEQTKDGTSSSSQKGFQDFOALGLKNTRYH 1487
Db	1807	-----TEKKEGEP-----VKGAEIEEQTKSNKRPIDEVELKANDE-- 1843
Qy	1488	GKLGISIRIPDGNELAKIKDASNNKKGEEKLLKSYDLFPKNYLNEVEKSKPKI 1539
Db	1844	-----DDTTENINESKKIKTED-----EBEKETDKV 1869

RESULT 8
ADS43855
ID ADS43855 standard; protein; 1875 AA.
XX

AA040883	61	NFTSDYQSVKALLNGKTFDPKSSSEPTDFVSKPDFLTNNGRVTLEIPKYYQVVISFSPSE	120
ID	AA040883	standard; protein; 2688 AA.	
XX	XX	AA040883;	
AC	XX	AA040883;	
DT	22-OCT-2001	(first entry)	
XX	XX	Human polypeptide SEQ ID NO 5814.	
DE	XX	Human; neurotropic; immunosuppressant; cytostatic; gene therapy; cancer;	
XX	XX	peripheral nervous system; neuropathy; central nervous system; CNS;	
KW	KW	Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;	
KW	KW	amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;	
KW	KW	chemokinetic; thrombolytic; drug screening; arthritis; inflammation;	
KW	KW	leukaemia.	
XX	XX	Homo sapiens.	
OS	XX	Homo sapiens.	
XX	XX	W0200153312-A1.	
PN	XX	26-JUL-2001.	
XX	XX	26-DEC-2000; 2000WO-US034263.	
XX	XX	23-DEC-1999; 99US-00471275.	
PR	21-JAN-2000; 2000US-00488725.		
PR	25-APR-2000; 2000US-00522317.		
PR	20-JUN-2000; 2000US-00598042.		
PR	19-JUL-2000; 2000US-00620312.		
PR	03-AUG-2000; 2000US-00653450.		
PR	14-SEP-2000; 2000US-00662191.		
PR	19-OCT-2000; 2000US-00693036.		
PR	29-NOV-2000; 2000US-00727344.		
XX	XX	(HYSE-) HYSEQ INC.	
PA	XX	Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;	
PI	XX	Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J, Zhao QA;	
PI	XX	Zhou P, Goodrich R, Drmanac RT;	
PI	XX	WPI; 2001-442253/47.	
DR	XX	N-PSDB; AA160039.	
XX	XX	Novel nucleic acids and polypeptides, useful for treating disorders such	
PT	XX	as central nervous system injuries.	
PT	XX	Example 2; SEQ ID NO 5814; 10078pp; English.	
PS	XX	The invention relates to human nucleic acids (AA157798-AA161369) and the	
XX	XX	encoded polypeptides (AA038642-AA042213) with neurotropic,	
CC	XX	immunosuppressant and cytostatic activity. The polynucleotides are useful	
CC	XX	in gene therapy. A composition containing a polypeptide or polynucleotide	
CC	XX	of the invention may be used to treat diseases of the peripheral nervous	
CC	XX	system, such as peripheral nervous injuries, peripheral neuropathy and	
CC	XX	localised neuropathies and central nervous system diseases, such as	
CC	XX	Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic	
CC	XX	lateral sclerosis, and Shy-Drager Syndrome. Other uses include the	
CC	XX	utilisation of the activities such as: Immune system suppression,	
CC	XX	Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic	
CC	XX	and thrombolytic activity, cancer diagnosis and therapy, drug screening,	
CC	XX	assays for receptor activity, arthritis and inflammation, leukaemia and	
CC	XX	C.N.S disorders. Note: The sequence data for this patent did not form	
CC	XX	part of the printed specification	
XX	XX	Sequence 2688 AA;	
SQ	XX	Query Match 3.1%; Score 302; DB 4; Length 2688;	
XX	XX	Best Local Similarity 19.5%; Pred. No. 4.3e-08;	
XX	XX	Matches 373; Conservative 320; Mismatches 689; Indels 530; Gaps 92;	
QY	1	MKNKSTLLATAAAIIGSTVFGVWGLASKVKYRGVWNPQTQGVISOLGLIDSVAFKPSIA 60	
DB	787	IQDKSELHITSEK---DKLFSEVHKESRV-----QGLLERIG-----KTKODLIA 830	

Db 1214 TERDHLRGYIREIB---ATGL---QTKEELKIAH-----IHUKHEQETIDELRRSVSEK 1261
Qy 537 SKAINNOGLEEDDNIITERIPENSPIYOQENAGLGASDPKPMIKDVQNRQRYLAKSQI 596
Db 1262 TAQIINTQDLEKSH---KLQEEIPVLHSEQL-----LPNVKVKSETQETMNELELL 1311
Qy 597 QELIKADYTKLAKLLNRYTHYINISRLKEQLFDVNPRIPS-SRDIKAKFVLDKTEKNK 655
Db 1312 TEOSTTKDSTTLARIEMER-----LRLNEKFQESQEBEIKSLTKERDNLTKIKALEV-K 1364
Qy 656 YWQIYSSASPVQNKWSLFGVYRYLLGLDPKQTHIE-LVKLGQKAGIQFQGENLPDSDFN 714
Db 1365 HDQL-----KEHIRETLAKIQESQSKQ-----EQLSLN 1391
Qy 715 LEDLKNIRIKTPLFSQDNFKLSLLDFNNYDGEIKAPBFGPLFLPKELRRNSSNGGS 774
Db 1392 MKEKDN--ETTKIVSEMEQFKPK-----DSALLRIEIM-LGLSKRLQESH-----1435
Qy 775 QNSNSPWEQEIISQFKQNLNSQDLAQFSTKIWE-----KIIGDENEPDQNNRIQYKLLKD 831
Db 1436 -----EMKSAKEKDDLQRLQEVLOSQSDQLKENIKETIVAKHLETEBELKVAHCCLE 1488
Qy 832 LOESWINKTRDNL-----YWTYLGDKLVKPKKNLEAK-----FROISNLQELL 875
Db 1489 -QETINELRVNLSEKETEISTQKQLEANDLQNKIQEIYKBEQNLTKQISEQVENV 1547
Qy 876 TAF-----YTSAAALNNWNYQDSGAKSTIIFBEIABLDPKVKEKGVADY 920
Db 1548 NELQKQKHEKAKDSALQIESKMLELTNRLQESQEBEIQIMIKEBEM-KRVQEAQIER 1606
Qy 921 YQLKHYAIGFDDNAGFNQEVNTRSSRTI-YLKTSGSKLEAD--TIDOLNQAQVKNAPL 977
Db 1607 DQLK-----ENTKEIVAKMESQEKYQFLKMTAVNETQKMCIEHLKQFETQKL 1658
Qy 978 GLQSFLYDTRFRFGVKLATSLAVQH-----KQEKTLPKKLNDGYTILHDKLKK 1028
Db 1659 NLEN--JETENIRITQILHENLEEMRSVTKERDRLRSVEETL--KVERD--QKENLRE 1711
Qy 1029 PVIQISSPEKWFEGKLNQSGQNVNVTSGSIIESPYFTNFOEDADLDQDQ---1084
Db 1712 TITRDLEKQEBELKIVHMLKHEH---QETIDKLRGIVSEKTNESNMQKDLHSNDALKAQ 1768
Qy 1085 -----QDSROGNSLNDQ-----AGLLQKQKAILL-----GNQFIQYQOQNDKEI 1126
Db 1769 DLKIQEBELIAHMLKQEQETIDKLRGIVSEKTDKLSNMQKDLNSNAKLQEKIQELKAN 1828
Qy 1127 EFETIINV-----EKVSLSFRVEFKLAKTLEDNGKTIRVLSDETMSLIVNTTIEKTP 1178
Db 1829 EHQUITLKVDNETQKQVSEH-----QLKKQIKDQSLTSLKLEIENLNL-AQELHENLE 1882
Qy 1179 EMSAVPEVFDT-KWVEQ-----YDPRTPLAATKTFVLKFKDQIPVDG 1219
Db 1883 EMKSVMERDNLNRVEETLKLERDQLKESLQETKARDLEIQBELKTARMLSKHEKETVD- 1941
Qy 1220 SGNISDKWLASIPLIVIHQOMLRSLSPVVKTRIEGLKTEQOQQOQQOQQOQQOQKQVAKRE 1279
Db 1942 --KLREK-----ISEKTIQSDIOK-----DLDKSKDELQKTIQELQKXELQLLRVK 1986
Qy 1280 EELETYNPKDEFNLTNPLTKAHLTLNLVNDPNYKIEDLKVTQNEAGDHQLAFSLRAN 1339
Db 1987 EDVNMSHKK--INEMEQKKQF-----FENY-----LCKCEMDNFQITKKLH-- 2026
Qy 1340 NIKRLMNTPTIFADYNPFYFNEDWRSIDKYLANKGN--VSSHQQOQAGNGQSGLIQRL 1397
Db 2027 --ESLEEIRIVAKE-----RDELRLIKESLSKMERDQFIATLREMIARDQ-----2069
Qy 1398 NKAIKPTFTPALIALKDRNNTNLNSYSDKIIIMPKPYLVERSIGVPWSTGLGYTGSEQ 1457
Db 2070 NHQVKPEK-----RLLSDDGOHLMESLREKCSRIKE--LLKR-----YSEMDHY-----2112
Qy 1458 TKDGTSSSSQKGPQDFQIALGL---KNTEYHG-----KLGLSIRIFDPGNEIAKTDA 1509
Db 2113 -----ECLNRLSLDLEKEIEPHRIMKKLVLSY-----VTKLEE 2148

RESULT 13
ABG06301

ID ABG06301 standard; protein; 2017 AA.

XX AC ABG06301;

XX DT 13-FEB-2002 (first entry)

XX DE Novel human diagnostic protein #6292.

XX KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
food supplement; medical imaging; diagnostic; genetic disorder.

XX OS Homo sapiens.

XX PN W0200175067-A2.

XX PD 11-OCT-2001.

XX PF 30-MAR-2001; 2001WO-US008631.

XX PR 31-MAR-2000; 2000US-00540217.

XX PR 23-AUG-2000; 2000US-00649167.

XX PA (HYSE-) HYSEQ INC.

XX PI Drmanac RT, Liu C, Tang YT;

XX DR WPI: 2001-639362/73.

XX DR N-PSDB; AAS70488.

XX PT New isolated polynucleotide and encoded polypeptides, useful in
diagnostics, forensics, gene mapping, identification of mutations
responsible for genetic disorders or other traits and to assess
biodiversity.

XX PS Claim 20; SEQ ID NO 36660; 103pp; English.

XX CC The invention relates to isolated polynucleotide (I) and polypeptide (II)
sequences. (I) is useful as hybridisation probes, polymerase chain
reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
and in recombinant production of (II). The polynucleotides are also used
in diagnostics as expressed sequence tags for identifying expressed
genes. (I) is useful in gene therapy techniques to restore normal
activity of (II) or to treat disease states involving (II). (II) is
useful for generating antibodies against it, detecting or quantitating a
polypeptide in tissue, as molecular weight markers and as a food
supplement. (II) and its binding partners are useful in medical imaging
of sites expressing (II). (I) and (II) are useful for treating disorders
involving aberrant protein expression or biological activity. The
polypeptide and polynucleotide sequences have application in
diagnostics, forensics, gene mapping, identification of mutations
responsible for genetic disorders or other traits to assess biodiversity
and to produce other types of data and products dependent on DNA and
amino acid sequences. ABG00010-ABG30377 represent novel human diagnostic

Db	1349	P-----YKFLNKEKD-----KFLSSYNIKSIDTIDINFANDVLGYKILSEKYKSLDLS	1399
Qy	1367	IDKYLNNK 1374	
Db	1400	IKKYINDK 1407	
RESULT 15			
ID	AAE29345		
XX	AAE29345	standard; protein; 1639 AA.	
XX	AAE29345;		
AC	AAE29345;		
XX	27-JAN-2003	(first entry)	
DT	XX		
XX	XX	Plasmodium falciparum merozoite surface protein-1 (MSP-1).	
DE	XX		
XX	XX	Band 3 polypeptide; malarial infection; drug resistance; vaccine;	
KW	XX	protozoacide; gene therapy.	
KW	XX		
XX	XX	Plasmodium falciparum.	
OS	XX		
XX	XX	WO200270542-A2.	
PN	XX		
XX	XX	12-SEP-2002.	
PD	XX		
PF	XX	01-MAR-2002; 2002WO-US006415.	
XX	XX		
PR	XX	02-MAR-2001; 2001US-0272930P.	
XX	XX		
PA	(SELI-) ST ELIZABETH'S MEDICAL CENT INC.		
PI	Chiahti AH, Oh SS, Liu D, Goel V;		
XX	XX		
DR	WPI; 2002-759814/82.		
DR	N-PSDB; AAD46980.		
XX	XX	New isolated Band 3 polypeptide which selectively binds to merozoite	
PT	PT	surface protein-1, useful for the prevention and treatment of malarial	
PT	PT	infection.	
XX	XX		
PS	Disclosure; Page 113-118; 163pp; English.		
XX	XX	The invention relates to an isolated Band 3 polypeptide that comprises	
CC	CC	any of 4 20 residue amino acid sequences, or their fragments that bind to	
CC	CC	an MSP-1 polypeptide or a polypeptide with any of 8 291-1331 base pair	
CC	CC	sequences. The methods and compositions of the present invention are	
CC	CC	useful for the prevention and treatment of malarial infection. The	
CC	CC	present invention, develops new and more improved methods based upon	
CC	CC	inhibiting the particular interactions between the malarial parasite and	
CC	CC	a cognate molecule present in the host and subsequently minimising	
CC	CC	harmful side effects and drug resistance that may be due to non-specific	
CC	CC	therapeutic approaches. The invention is useful in gene therapy. The	
CC	CC	present invention also provides a vaccine for malaria. The present	
CC	CC	sequence is Plasmodium falciparum merozoite surface protein-1 (MSP-1)	
XX	XX		
SQ	Sequence 1639 AA;		
Query Match			
Best Local Similarity 3.0%; Score 295; DB 5; Length 1639;			
Matches 282; Conservative 204; Mismatches 478; Indels 484; Gaps 68;			
Qy	203	LNQFNQDELWVYLEKFFDLEALKANIRLQADFSPEKGNLVDPPFVYSFIRPQKQKWA	262
Db	168	IDGYEINELLYKLNFFDILLRAKLN-----DVCANDYCOIIPFNLKIRA	211
Qy	263	SDLNQOQKTVRLYLRTEPSPQAKTLKDYKYKDETFLLSSIDLKASNGTSLFANENDLQD	322
Db	212	NELDVLKGLV-----FGYRKPL-----DNIKDN	234
Qy	323	LDVDDLDVSDYFGGQSETITNSQVKPVPASERSLKDVRKFKDQOKPRIEK-----FSLY	378
Db	235	VG-----KMEDYIKNNKKTININEL--IESKKTIDKNKNATKEEKKLYQAQYDLSIY	288

Qy	379	-----EYDALSPYSOLQELVSKPNSIKDLVNAI-----	406
Db	289	NKOLEBAHNLI SVLEKRIDTLKKNENIKELDKINEIKPPPPANGTPTNLLDKNKKIE	348
Qy	407	-----LARNLRFSLGKYNFLFDDLASHLDYVFLVSKAKIKOSSITKKLFLFELPKIS	458
Db	349	EHEKEIKEIAKTIKFI--DSLFTD-PLLELYL--REKNKN-----IDISAKVE	393
Qy	459	LKSSILGDQEPNIKTILFEKEVTFKLNFDRDVEIEKAFGLLYPGVNELEQARK-----AQR	514
Db	394	TKESTEPNEYPN-----GVTYPLS-----YNDINNALNEINSFGDLINP	432
Qy	515	ASPEKEKSKGLKEFQOKEENSAINNOEGLBEDDNITERLPENSPIOYQOENAGLGAS	574
Db	433	FDYKEFSKNIYTD-----NERKFFINE--IKEKIKIEKKIESDKKSYE-----	475
Qy	575	PDKPYMIKDVQONORYYLAQSIQIELIKAK-----DYTKLAKLLSNRUTYINISLRKQLP	629
Db	476	-DRSKSLNDITKEYEKL--LNEIYDSKFNNNIDLTFEKMGMKRYSYKV-----EKL	525
Qy	630	DVNPRIEPRDIEKAFVLDK-TEKNKYWQIYSSASPVFQNKWSLFGYVRYVLGL--DPK	686
Db	526	HHN-----TFASYENSNGNLEKLTAKALYMSDYSURNIVVEKE--LKYIKNLISKIENEI	578
Qy	687	QTIHELVLGOKAGLOFEGY---ENLPSDFNLE--DLKNIRIKTPLFSQK-DNFKLSLL	739
Db	579	ETLVENIKDBEQ--LFEKKITKDNKPKDEKILEVSDIVKVQVKVLLMKNKIDELKKTQL	636
Qy	740	DFNN-----YYDGIKAPFGLPLPLPK-----ELRNSNSG	772
Db	637	ILKNVELKHNIHVPNSYKQENKQEPYLI VLKKEIDKLVFMPKVESLINEEKNIKTEG	696
Qy	773	GSQNSPFWOEIISQF-----	789
Db	697	QSDNSEPSTEGETTGQATTKPGQOAGSALGDSVQAQAEQKQAPPPVPVPEAKAQVP	756
Qy	790	-----KQNLNSQDLAQFSTKIWE--KIIDGENEFQDNRLQYKLLKD	831
Db	757	TPPAPVNKNTENVSKLDYL-----EKLYEFLNTSYICHKYLIVSHSTWNEKILKQYKITKE	812
Qy	832	LQESWINKTRDNLWYTLGDKLVKPKKNLEAKFQI-----SNIOELLTAPYTSAAUNNN	887
Db	813	-EESKLSK-DPL-----DLLEFNQNNIIPVMTSMFDSLNNLSQLFMEIYEMVNCNL	863
Qy	888	WNYQDSGAKSTIIFERIAELDPKVEKGVADVQLK-----	924
Db	864	YKLKDNKIKN--LLEBAKKYSTSVKTLSSSMQPLSLTPQDKPEVSANDDTSHSTNLN	921
Qy	925	---PHYAIGFDDNAGKFNQEVIRSSRTIYLTSGKS-----KLEADTIDQL	968
Db	922	SLKLFENILSGKKNKIYQELIGQKSENFEYKILKDSDTFYNESFTFVKSKADDINSL	981
Qy	969	NQAVKNAPLGLQSFYLDTERFGVFOKATSLAVQHKOK-EKTLPKLNNDGYTLIHDKLK	1027
Db	982	NDESKRKKL-----EEDINKLKTQLSFLDLYNKYKLERLFDKKTGTGKYKM---QIK	1033
Qy	1028	KPVIQIOTSSSPKDFEGLKNQOSQNV--NVSTFGSIIIESPYFTSFQSDADLDQDQ	1085
Db	1034	KLTL-----LKEQLESKLNLANPCHVLQNFVSF-----FNKCKEAEI-----	1071
Qy	1086	DDSRGNNSLDNOEAGLLKQKLAILLGNQFIQYQOQND---KEIEFEIINVEK--VSELS	1140
Db	1072	---ASTENLENTKI-LLKHVKGL-----VKYNGESSPLKLTSEESIQTEDNYASLEN	1121
Qy	1141	FRVEFKLAKLTEDN---GKTRIVLSDETSLI-----VNTT	1173
Db	1122	FKVLSKLSGKLKDLNLNLEKSKLVSLSGLHLLAELKEVKNKNTGNSPENNTDVNA	1181
Qy	1174	TK-----TPMSAVPEVF-----DTKWVEQYDPRTP---LAAKTFVLKPKDQ-----	1214
Db	1182	LESYKKFLPEGTDVATVVSSEGSdT--LEQSQPKPASTHVGAESNTITTSQNVDDEVD	1239


```
Db 1134 ENFKVSKLEGLKDNLNLEKKLSYLSRGLHLHIAELKEVKNKNTGNSPNSVNTDVA 1193
Qy 1172 TTIEK----TPMSAVPEVF-----DTKWVEQYDPRTPLAAKTKFVLKFKDQIPVDGSGN 1222
Db 1194 NALSYKKFLPEGTVDVATVWSESGDT--LEQSQPKKPASTHVG-----AESNTITTSQN 1246
Qy 1223 ISDKWLASIPLVI-----HQQMLRLSPVVKT 1248
Db 1247 VDDEVDVIVLIFGESEEDYDVGQVVTGEAVTTSVIDNLSKIENEYEVLYLKPLAGV 1306
Qy 1249 IREGLKTEQOQQOQQOQQOQKQKAVRKEELET-YNPKDEF-NIL-NPLTKAHLTL 1305
Db 1307 YRSL-----KKQLENNVMTFNVVVKDILNSRFNKEFNKVLSDLIPIYKDLTS 1355
Qy 1306 SNLVNDPNYKIEDLVKIKNAGHQHAFSLRANNIKRLMNTPTITPADYNPPFY----- 1359
Db 1356 SNYVVKDP-----YKFLNKEKRD---KFLSSVNYIKDSIDTDINFANDVLGYKILSEK 1406
Qy 1360 YNEDWRSIDKYLNNKGNVSSHQQQAAGNQGSLIQRLNKNIKPTFTTPALIALKDRNN- 1418
Db 1407 YKSLDLSIKKYINDK-----QGEN-----EKYLPFL-----NNI 1435
Qy 1419 -TLSNYSDKI-----IMIKPK---YLVRSIGVPMSTGLDGYIGSEQTKDGTSSSSQK 1470
Db 1436 ETLKYTVNDKIDLVFIHLEAKVLYTEKS-----NVEVK----- 1470
Qy 1471 FDQDFIQALGNKTEYHGLGLSIRIFDPGNELAKIKDASKKGEKLLKSY----- 1522
Db 1471 -----IKELNLYKTIQKADPKNNFVGIADLSFDYNNHNLTKFLSTGMVFE 1520
Qy 1523 DLFKNYL-NEYEKSPKAKGTN-----IHPDQKE-----YPNPNQK 1559
Db 1521 NLLKSVLSNLDWKLARYVGHFTPMRKKTWIOSSGCFRHLDERECKLLNYKQEGSK 1580
Qy 1560 LPEN 1563
Db 1581 CVEN 1584

RESULT 17
AAR05591
ID AAR05591 standard; protein; 896 AA.
XX AAR05591;
AC AAR05591;
XX
XX
DT 25-MAR-2003 (revised)
DT 07-AUG-1990 (first entry)
XX
DE The 36 kD antigen of Mycoplasma hyopneumonia.
XX
XX Recombinant Mycoplasma hyopneumoniae antigens; mycoplasma pneumonia;
KW vaccine; antibody; ds.
XX
OS Synthetic.
XX
FH Location/Qualifiers
FT 78..78
FT /label= Coded by TGA nonsense codon.
FT 292..292
FT /label= Coded by TGA nonsense codon.
FT 298..298
FT /label= Coded by TAA nonsense codon.
FT 366..366
FT /label= Coded by TAA nonsense codon.
FT 370..370
FT /label= Coded by TAA nonsense codon.
FT 377..377
FT /label= Coded by TGA nonsense codon.
FT 414..414
FT /label= Coded by TAG nonsense codon.
FT 420..420
FT /label= Coded by TAA nonsense codon.
FT 527..528
```

```
FT Region
FT 534..534
FT /label= Coded by TAA nonsense codon.
FT 539..539
FT /label= Coded by TAA nonsense codon.
FT 549..549
FT /label= Coded by TAA nonsense codon.
FT 559..559
FT /label= Coded by TAA nonsense codon.
FT 664..664
FT /label= Coded by TAA nonsense codon.
FT 676..676
FT /label= Coded by TGA nonsense codon.
FT 677..677
FT /label= Coded by TGA nonsense codon.
FT 725..726
FT /label= Undefined region of 50 units long.
FT 729..729
FT /label= Coded by TGA nonsense codon.
FT 734..734
FT /label= Coded by TAA nonsense codon.
FT 743..743
FT /label= Coded by TGA nonsense codon.
FT 744..744
FT /label= Coded by TAA nonsense codon.
FT 756..756
FT /label= Coded by TAA nonsense codon.
FT 788..788
FT /label= Coded by TAA nonsense codon
FT 790..790
FT /label= Coded by TGA nonsense codon.
XX
XX EP359919-A.
XX
XX 28-MAR-1990.
XX
XX 28-JUN-1989; 89EP-00111748.
XX
XX 29-JUN-1988; 88US-00213248.
PR 07-APR-1989; 89US-00334586.
PR 21-APR-1989; 89US-00341968.
XX
XX (MLTE-) ML TECHN VENTURES L.
XX
XX Faulds DH, Brooks E, Andrews WH, Lory C;
XX
XX WPI; 1990-092610/13.
DR N-PSDB; AAQ03634.
XX
XX Recombinant Mycoplasma hyopneumoniae antigens - used in vaccines against
XX mycoplasma pneumonia and in assays for detection or determ. of
XX antibody.
XX
XX Disclosure; Fig 34; 54pp; English.
XX
XX An expression vehicle containing the sequence is capable of eliciting an
XX antibody which recognises an epitope of an M. hyo antigen. The
XX recombinant proteins can be used in a vaccine for protection against
XX Mycoplasma pneumonia, partic. in swine, br in an assay for detection of
XX antibodies to M.hyo. All X's in the sequence that are not coded by
XX nonsense codons are blanks in the specification. See also AAQ03629-
XX 003634, AAQ04959-Q04963 and AAQ02474. (Updated on 25-MAR-2003 to correct
XX PR field.) (Updated on 25-MAR-2003 to correct PA field.)
XX
XX Sequence 896 AA;
SQ
Query Match 3.0%; Score 291.5; DB 2; Length 896;
Best Local Similarity 21.8%; Pred. No. 3.4e-08;
Matches 131; Conservative 109; Mismatches 240; Indels 121; Gaps 22;
Qy 1 MNKKSTLLLATAAAIIGSTVFGTVGLSKVKVGVNPTQGVISQLGLDSVAFKPSIA 60
Db 1 MSKSKTFKIGLTAGIVGLVGLTGLSLAKVRSBSPRKIANDFAAKVSTLAFSPYAF 60
```


Db 1001 -----DLVSNFV-VLNYGVNEYDFNLDDLKH 1027
 Qy 429 LDYVFLVSKAKIQSSITKKLFIELPIKISLSKSIILGDQEPNITKLFKEVETIFKLNFRD 488
 Db 1028 RIYSF-----KKLIYEPNQDKYFSFLNQKIT-----NTEKT----- 1061
 Qy 489 VEIEKAGLLYPGVNEBLEQAKQARASPEKESKGLK-----EFSQO-----K 533
 Db 1062 -----ASVNEQLTIDTKLVKRP-DSNLKLNKLQINDPNDPFLQDSILETFH 1109
 Qy 534 EENSATNNQGG-LEEDNITERLPENSPYOQOENAG-LGASDPKPYMIKDVQNOYYL 591
 Db 1110 DEKDKTHNVIGKINVDAN-----NNTLEFSVENTNTFKIQNHXYIVDNIN-----YA 1159
 Qy 592 AKSQIOELIKAKYTKLAKLSSNRHT-----YNIS-----LRLKEQLF-----D 630
 Db 1160 TKNKIQP-----ANASNNNSKNISYDASSNPSKILSFTNELLFVNINNVNPNNTN 1209
 Qy 631 VNP-----RIPSRDIEKAFV-----LDKTEKNKYQIYSSASPVF----- 667
 Db 1210 LNPATASIDVELKSSONLLKQYLRALYIDNNHQ-KIWSDYASVNVNVDLAHINLNFNLTP 1268
 Qy 668 QNKSLFGYVYLLGLDPKQTIHEL--VKLGOKAGLOFEGYENLPSPDNLEDLKNIIRKT 725
 Db 1269 NRKYNFAGIYF-----NDQONNDETQKKIFVKESKTYIDFSTLASTTQLLEFKTHNV-- 1323
 Qy 726 PLFSQONFKLSLLDFNNYDGEIKAPEFGPLFLPKELRRNSNGSGSONSPWQEI 785
 Db 1324 -----DEQDN--YDFNINDDQVLEPGMLVELCF----- 1351
 Qy 786 ISQKQNLNODQ-LAQFSTKIWEKIIGDENEDQNNRLQYKLLKOLQESWINKTRDNL 844
 Db 1352 -----ENLVDKNQPLVKTQTTLVKK-----DEHSFSASGVI-----TNLLADHT 1390
 Qy 845 YWYVLGDKLVKPK-----NNLEAKPQISNLQELLTAFYTSAAALSNNMNYQDSG 895
 Db 1391 Y-RLTSVGLKQKPLQANVNNNNNNNEILLNIEDNQTIHTL----- 1432
 Qy 896 AKSTIIFEEIABL-DPKVKVGVADVYQKHYAIGFDNAGKQVQEVIRSSRTIYLT 954
 Db 1433 --SKSIVNEISNLADAYPKNNNG--YDVKEN--INIIKNNKLNKYVNVFEDNEHQL 1486
 Qy 955 SKSKLEADTIDQLNAQVAPLGLQSF-----YLDTERGVFQKLAT 997
 Db 1487 ISTNDLLVNLKDOTNL-----ISLQNFSPNLKPNHLYRLKVVYGEQNF----- 1533
 Qy 998 SLAVQHKQKETPLKLANNDGYTLIHDKLPVPIQISSPEKDWFEGL-----N 1048
 Db 1534 --AINEQKNILALNPSLVNSFSSTPAKIK-----VSKNAIDVWQNALIKLILDDSDN 1585
 Qy 1049 Q--NGSQSNVNSTFG--SIESPYFSTNFQEDADLDQDQDSDSQGNSLONQAGLLK 1104
 Db 1586 QLSHGEININRIKGTQNIISTPATIS-----DQDKRYTKCMATNLI-----AGLDY 1633
 Qy 1105 QKLAILLGNQFTQ-----YQOQNDKEIFEI-I 1131
 Db 1634 EIVSVTIKQKTNVSPFIPELSPGPIGYTLAPVKITSMDPVYHENKNTADKLNL 1693
 Qy 1132 NYEKVSELSFRVEFKLAKTEONGKTIIRVLSDETMSLIWNTTIKTPMSAVPEVFTKW 1191
 Db 1694 RIENIGASLNFNDIKFIFKKNKGQKQISFIHKVT-----SANDANYEWF 1738
 Qy 1192 VEQYDPRTPLAARTKFLVLPKD--QIPVDGSGNISDKW-----LASIPLVHQ 1238
 Db 1739 KOLLRRNREYTLERVVY-LANKDFNQSNGVSESYI-DLWIYENLNKTFKLLPTKPLGI--- 1793
 Qy 1239 MLRLSPVVTIRELGLKTEQO-----OQOQOQOQOQOQKAVRKEELETYNPK 1288
 Db 1794 ---IGAPIKEISDNGAKVQLKFAINDFDVLKENQTFKNIQPNENNANLNEISEHECK 1850
 Qy 1289 DEFNINPLTKAHLRLTSLNVNDPN--YKIEDLVKINE---AGDHOLAFS---LRANN 1340
 Db 1851 VE-----IIDGQKFFVANLNKIKVNEYKVYKIIFDENQDVNGVYKINFKNDYKEPN 1904

Qy 1341 I--KRLMNTPIPADYNPEFY--YNEDWRSID-----KYL 1371
 Db 1905 VYDASVNTTQYVFTTKFAIASFSNNLTDVDVANKQNTSINLDSRVETIQGVHFPAKVI 1964
 Qy 1372 -----NNKGNVSSHQOAGGNGSGLIORL----- 1397
 Db 1965 SINDRVVNTTIPAPTINGNKNKNIALNPEL---NONQLISNRLYTFALAYYSKDVNDV 2020
 Qy 1398 -----NKNIKPTFP---TPALIALK-----DRNNTLS---NYSKIMIKPKYLV 1437
 Db 2021 ENHANVVKNNVNPOTISTKPSSTYVDLKAQNADEKNTLSLLLSNDQOIFEDKNNHLK 2080
 Qy 1438 ERSIGVPMSTGLDGYLGSQTKDGTSSSQOKGPDQDFQALGLK-NTEYHGK----- 1489
 Db 2081 IAKISIDELDAHQIINS--TTHDYDITLKEKNENWLLKTQLINLKNPTKYRVKVKWFTSK 2139
 Qy 1490 -----LGLSIRIFDPG-----NELAKIKDASNKK 1513
 Db 2140 PSDTIYGVNKNIVYPINHNNTNIDLKTLEKSTLNSVKPTKTNFNKNQESIKLMVGFNKT 2199
 Qy 1514 G---BEKLLKSYDLFKNYLNEYEKSPKIAKGTNTNIHPDOKEYPNPNQKLPENYLNVLN 1570
 Db 2200 GSSLENKYAKL--VYKDNNNQIESDEVLNLSDDTTTKPFNKFSFNN---NOTSLVAN 2252
 Qy 1571 QPW---KVTLYNSDDF-----ITNLF-----VEPEGSDRGSGTKLKQVIQOVNN 1612
 Db 2253 RDEFEVKLIISDTADFQNTANTLDTNFKNDKDAKFSIEP-----TPISVNNVQSGSNT 2306
 Qy 1613 NYADWGSAYLTFWY---DKNIITN-----ITANIADVFVKDKELEDN 1656
 Db 2307 ---YDEIHLSPDYDDQDHLVDNDQITITIRYKKGQAMTSKAGEVSVKNHKINVTLD 2421
 Qy 1634 ---QPNV-----ITANIADVFVKDKELEDN 1656
 Db 2363 LTPNTTVEIGAISTHVSFNVKTSPIQYNETTKOLEKLDLTKVAHNFIKDPK-LASN 2421
 Qy 1657 TKLI-----APNITQWPNISGSKEKFKYKPTVFF 1685
 Db 2422 SKLVGVYECADENKVFVGLHKEGHELEGYEIYAKFVKTNP-----KPAIEI 2472
 Qy 1686 GNWENENSSNNSOATPTWEKIREGALQALXKSFDQKTRFTVLTNAPLPMKYGLGF 1745
 Db 2473 KSSNNINSN-----QVYEFELLSQO-----ITPNHQVSELSICAEN 2511
 Qy 1746 QNGENFTQDWRVLFQNDNDQIAALRV---QEQDRP-----EKSEDKDKQWIKFKVVI 1797
 Db 2512 KGANNEIKKO-----LNENHLSLALGLAPNKMVQKPTITITIKVDNQSVSAHVKIIND 2566
 Qy 1798 PEEMFNSG-----NIRFVGVMQIQGPNTLWLPVINSVVIYD-----FYRGTDGSDNVA 1845
 Db 2567 PDGLTNSDSCSKELLNKKYALFNQTNANL---INNITIKYDQONKQFY---CEPDID 2618
 Qy 1846 NLNV-APQVQKTIATFTNNAFNNVKEFN 1872
 Db 2619 NLTLNQDYGIFEISFANKPIHAAPAKIN 2646

RESULT 19
 ADS43650
 ID ADS43650 standard; protein; 1679 AA.
 XX ADS43650;
 AC ADS43650;
 XX
 DT 02-DEC-2004 (first entry)
 XX
 DE Bacterial polypeptide #22080.

Recombinant DNA construct; transformed plant; improved plant property;
 cold tolerance; heat tolerance; drought tolerance; herbicide; osmotic;
 pathogen tolerance; pest tolerance; plant disease resistance;
 cell cycle pathway modification; plant growth regulator;
 homologous recombination; seed oil yield; protein yield; carbohydrate;

nitrogen; phosphorus; photosynthesis; lignin; galactomannan;
bacterial polypeptide.

Bacteria.
US2003233675-A1.
18-DEC-2003.
20-FEB-2003; 2003US-00369493.
21-FEB-2002; 2002US-0360039P.
(CAOY/) CAO Y.
(HINK/) HINKLE G J.
(SLAT/) SLATER S C.
(CHEN/) CHEN X.
(GOLD/) GOLDMAN B S.
Cao Y, Hinkle GJ, Slater SC, Chen X, Goldman BS;
WPI; 2004-061375/06.
New recombinant DNA construct comprising a promoter positioned to provide
for expression of a polynucleotide encoding a polypeptide from a
microbial source, useful for producing plants with improved properties.
Claim 1; SEQ ID NO 22080; 122pp; English.
The invention relates to a recombinant DNA construct comprising a
promoter functional in a plant cell, where the promoter is positioned to
provide for expression of a polynucleotide encoding a polypeptide from a
microbial source. The invention also relates to a transformed plant
comprising the recombinant DNA construct and a method of producing a
transformed plant having an improved property. The plant is a crop plant
such as maize or soybean. The method of producing a transformed plant
having an improved property comprises transforming a plant with the
recombinant DNA construct and growing the transformed plant, where the
polynucleotide or polypeptide is useful for improving plant properties.
The recombinant DNA construct is useful for producing plants with
improved plant properties, e.g. improved cold, heat or drought tolerance,
increased resistance to plant disease, better growth rate by modification
of the cell cycle pathway with plant growth regulators, increased rate of
homologous recombination, modified seed oil or protein yield and/or
content, improved yield by modification of carbohydrate, nitrogen or
phosphorus use and/or uptake, by modification of photosynthesis or by
providing improved plant growth and development under at least one stress
condition, improved lignin production or improved galactomannan
production. This sequence represents a bacterial polypeptide used in the
scope of the invention. Note: The sequence data for this patent did not
form part of the printed specification but was obtained in electronic
format from USPTO at seqdata.uspto.gov/sequence.html.
Sequence 1679 AA;
Query Match 3.0%; Score 290.5; DB 8; Length 1679;
Best Local Similarity 19.7%; Pred. No. 1e-07;
Matches 383; Conservative 312; Mismatches 684; Indels 567; Gaps 97;
QY 63 TSDYQSVKALLNGKTFDPKSSFFDFVSKFDFLTNGRTVLEIPKYQVWISFSPF-D 121
DB 69 SSSQKNKAEELNGK-DQLNEERSYRREIDAL-----KKQLHVSHEAMREYN 116
QY 122 DKERFRLGHKLEKLEBGN-----IAQSATKFIYLLPLDMPKALGOVSY-- 166
DB 117 DEKRVKEEYDIQSRDQNDLNDLKENKLLRRKLMENILQRCCKNALSLLQKYDT 176
QY 167 -IVDKNF-----NNLIHPLSNFSAQSIKPLALPRSS---DFTAKLQNF--NNOELAVY-- 215
DB 177 SVOKEIMLQSKKLIBELKSSFSKTLTB-EVTKSSHVENLEBKLYMQMSVESVFTYNK 235
QY 216 -----LEKFFDLKALKANIRLOTADSFSEKGNLVDPPVYSFIRNPQKQWA 262

Db 236 FILNQNKQLSQSVBEKVLEMKNLKDTASVEKAFFSKB-----MTIQKNM 279
QY 263 SDLNQDQKTVRLYLRTFSPQAKTILKQYKXDETFPLSSIDLKASNGTSLFANENDLKQ 322
Db 280 NDLLRSQLT---SLEKDCSLRAIEKNDNSCRNPEHTDVIDELIDTKLREKSKNCSQ-R 335
QY 323 LDVDDLVDYFGGSETITTSNQKVPVPAERSLKORVKFKDQKQKPKRIEKFSL---VE 379
Db 336 LQNIWMDCTK---BEEATMTTSVSPVTG--KLFSDIKVLKRQLIKERNQKFLQNLQE 389
QY 380 YDALSFPYQQLQELVSKPNSIKDLVNATLARNLRFSLGKYNFLFDDLASHLYVFLVSKAK 439
Db 390 DFILLEHKTPELISFKERTKSLH-----ELKRTELELTVSLTK 430
QY 440 IKQS-SITKKFLFELPIKISLAKSSILGDOEPNIKTILFEKEVTFKLDNFRDVEIEKAFGLL 498
Db 431 RKQEREIT-----SLRQKING-CEANIHSLVKQ---RLDLARQVKL-----LL 469
QY 499 YPGVNEELEQARKAQRASPEKEKSKGLKBFQSQEKSAINNOEGLEEDDNITERLPE 558
Db 470 ---LNTSAIQETASPLSQDELISLRKILESSNIVNENDSQAI-----ITERLVE 515
QY 559 NSPI-QYQENAGL-----GASPDKPYMIKDVONQRYLLAKSQIQEL--IK 601
Db 574 AKQETRIILLERDSYKLLASTEENKANTNSVTSMEAREKKIRELE-AELSSTKVEN- 631
QY 655 KYWQIYSSASPVFNKWSLFGYRYLLGDPKQTHLVLKQKAGLQEGEYENLPSDPN 714
Db 632 -----SAIQN-----LRKELLYKKSQCKKT-----T 655
QY 715 LEDLKNIRIKTFLFSQKONFLSLDFFNNYDGEIKAPEFGLPLPLPKELRRNSSNGS 774
Db 656 LEDFENFK---GLAKEKERMLEBEAID---HLKAELEKQKSWPVSIVHVEKRASTELSQS 709
QY 775 QNSNSPWEQIISQFQDNLS---NQDQLAQSTKIWEKIIGDNEFDDQNNRLQYKLLKD 831
Db 710 RIKIKSLEYE-LSLKKETASPIPTKESITR-----DFEQCKEKKELQMR 754
QY 832 IQESWINTKRDNLWYTYLGDULKVVKPK---NNLEAKFQIISNLQELLTAFYTSAAULNN- 887
Db 755 LKESBISHNENKMFDSKSGQYKAKIKELNLE---RLRSLQSKIQIESIRSCKSQ 811
QY 888 --W--NYQDSCAKSTIIFEEIAELDPKVKKGADVYQL-----KPHYAIGFDDNAG 936
Db 812 LKWAQNTIDDTBMKWSLTELUNKETTI-BKLSSEIENLDKELKTKFYQYKF-LDQNSD 869
QY 937 KFNQE-VIRSSSRTIYK-TSGKSKLEADTIDQLNQAVKNAPLQSLQSYLDFTEREGVFOK 994
Db 870 ASTLEPTLRKELEQIQVLKQDANSQIA--YEEIISSENENALIELKNELAKTK-----EN 922
QY 995 LATSIAVQHKQK---EKLTPKLLNDG-YTLIHDKLKPKVPIQIISSEPKQWFEGLKNQ 1050
Db 923 YDAKIELEKKEKWARDEEDLSRLRGLGETRALQPLKEGALHFVQOS-----EKLRENV 976
QY 1051 GOSQNV--NVSTFGSIIIE-----SPYFSTNFQSDAD-----LDQDQDSDSRGNNNS 1094
Db 977 ERIQMIKIEKQSTIVQLCKKKEWSQYOST-MKENKDLSELVIRLEKDAAD----- 1027
QY 1095 LDNQBAGLLKQKAILLGNQFIQYQQN--DKIEIFE---IINVEKVSSELSFRVE----- 1144
Db 1028 ---CQAEITKTSLSYSAQDLDDKHERKWMEEKADVERELISNIEQTE--SLRVENSVL 1082
QY 1145 FKAKTLEDNG-----KTRIVLS-----DETMSLIVN-- 1171
Db 1083 EKVDVDTAANNQDKOHLKLVSLFSLNRHERNSLETKLTTCKRELAFVKQKNSLEKTINDL 1142
QY 1172 ---TTIEKTPMSAVPEVFD-----TKWVEQYDPRTPLAAKTKFVLKFKDQIPVDGSGN 1222

Db 1143 QRTOTLSEKEQYQSAV--IIDEFKDITKEVTQV-----ILKENNALQKSLKN 1189
Qy 1223 ISDKWLASIFLVIHQM-----LRLS-PVVKTIRELGL-----KTEQOQOQOQOQ 1266
Db 1190 VTEK-----NREIYKQLNDRQEBISRLQDLIOQKEQVINSKNILVYSEMEQCKORYQ 1244
Qy 1267 QOQPOKAVRKEBELTYNPKDEFNLPLTKAHLRLTLNVLNPNQNYKIEDLVKIKNE 1326
Db 1245 DLSQOQKDAQKCO-----IEKLTN-----EISDLK----- 1269
Qy 1327 AGDHOLAFSLRANNIKRLMNTPIITFADYNPPFYNNEDWRSDIDKYLNNKGN---VSSHQOQ 1383
Db 1270 ---GKLSAENAN-----AD-----LENKFNRLKQAHEKL 1297
Qy 1384 AAGNQSGGLQRLNKNIKPFTFPALIAKORNTNLSNYSOKIIMIKPKYLVERSIGV 1443
Db 1298 DASKQQAALTNELNE-----LKAIDKLEQDLHPENAKVI----- 1333
Qy 1444 PWSGLDGYIGSEQTKDGTSSSQKGFQDQDFIQALGLKNTYHGLGLSIRIFDPCN-- 1501
Db 1334 ---DLDTKLKAHELQSEDVSRDHEKOTYRTLMEEI-----ESLKKELQIPKTANSS 1381
Qy 1502 -----ELAKIDASNKKGBEKLKSYDLFKNYLNEYEKSPKIAKWTNIHP 1548
Db 1382 SDAFEKLVNMEKEKDRIIDERTKEFEKQLOETLNKTSSEAEYSKDIEILKKEMLEKE 1441
Qy 1549 DQ-----KEYPNQKLP--ENYLNLV-----LNQPKVTLNNSDFITNLVFEPE 1592
Db 1442 DETLRRIKEAENLKRIRLPSERIQKISKRKEELEBEPKXKENAGSLTFL----- 1496
Qy 1593 GSDRGSGTKLQVQKQVNNYADWGSAYLTFWDKNIITNQPVIITANIA-DVFIKDVK 1651
Db 1497 -DNKSGSDEAEELWNSPKGNSRPSAVAFINQKRL---KPOEQLKNVKNQDVSNDSQ 1552
Qy 1652 ELDNTKLIAPNITOWPNINIGSKEKFKYKPTVFFGN-WENENSSNNSQAQPTTWEKIREG 1710
Db 1553 SMVTNKE---NNIVD---SSAAGNKA---IPTFSFGKPPFSSNTS----- 1598
Qy 1711 FALQALKSSFPQKTRTFVLTTNAPLPLWKYGP-LGFQNGPNPKTQDWRLVFPQNDNQIAA 1769
Db 1589 -SLOSFQNPFTASQSN--INTNAPLRLTNIQIEAVAKAAINFSN-----VTDLTNNSTDG 1640
Qy 1770 LRVQE-----QDRPEKS--SEDKQKQK 1789
Db 1641 AKITEIGTSKRPIESGTSDDPTTKK 1666

RESULT 20
AAB18171
ID AAB18171 standard; protein; 1979 AA.
XX AC AAB18171;
XX AC AAB18171;
XX DT 07-NOV-2000 (first entry)
XX DE Plasmodium falciparum chromosome 2 related protein SEQ ID NO:28.
XX DE Plasmodium falciparum; chromosome 2; human malaria parasite; vaccine;
XX KW antimalarial; malaria; protozoacide; infection; insecticide.
XX OS Plasmodium falciparum.
XX PN WO200025728-A2.
XX PD 11-MAY-2000.
XX PF 05-NOV-1999; 99WO-US026796.
XX PR 05-NOV-1998; 98US-0107131P.
XX PA (HOFF/) HOFFMAN S.
XX PA (CARU/) CARUCCI D.
XX PA (GARD/) GARDNER M.

PA (VENT/) VENTER J C.
XX Hoffman S, Carucci D, Gardner M, Venter JC;
XX WPI; 2000-365347/31.
XX PT Plasmodium falciparum, useful as antimalarial vaccines and in the
XX diagnosis of P.falciparum infection.
XX PS Disclosure; Page 70-75; 577pp; English.
XX CC The present invention describes proteins and their fragments (I) encoded
XX by chromosome 2 of the human malarial parasite, Plasmodium falciparum.
XX Also described are: (I) nucleotide sequences (II) encoding (I); and (2)
XX vaccines against P. falciparum infection comprising (I) or (II). (I) and
XX (II) are useful for the development of vaccines against P. falciparum
XX infection. (I) and polyclonal antisera or a monoclonal antibody raised to
XX immunogens comprising the sequences of (I), are useful in the detection
XX of infection with P. falciparum. Furthermore, (I) (especially when they
XX are rifins or secreted or membrane proteins) can aid the identification
XX of drugs to treat or prevent P. falciparum infection, or they can be used
XX to identify drug resistance in P. falciparum. Sequencing of the
XX Plasmodium chromosome 2 and the subsequent identification of parasite biology,
XX encoded by it will help to expand our understanding of parasite biology,
XX a process hampered by the complexity of the parasitic lifecycle, and
XX provide new targets for vaccine and drug development. Parasite resistance
XX to drugs and mosquito resistance to insecticides have led to a resurgence
XX of malaria in many parts of the world, and there is a pressing need for
XX vaccines and new drugs. AAA70078 to AAA70287 and AAB18144 to AAB18352
XX represent nucleotide and protein sequences given in the present
XX invention, but which are not specifically mentioned within the
XX specification
SQ Sequence 1979 AA;
Query Match 3.0%; Score 290.5; DB 3; Length 1979;
Best Local Similarity 17.1%; Pred. No. 1.3e-07;
Matches 293; Conservative 301; Mismatches 559; Indels 563; Gaps 73;
Qy 248 VYGFIRNPQKQWASDLNODQTVRLYLRTESPOAKTILKYKYKDEFLSIDLKAS 307
Db 2 VFTTF-KNKKKKKQSSD-----KVSKESENEEDNENNEKREKSDSWYKIIETKKG 51
Qy 308 NGTSLFANENDLKDQDLVDLVDYFGGQGEITTSNSQVQKVPASERSLDRVKFKKQ 367
Db 52 SKTK-YKNDNSLDNINEDIIINNNNNNNNDNNNDNN-----NNDNNNDNNNDNNE 104
Qy 368 QKPRIEKFSLYEYDALSFSQLOELVSKPNSEIKDLVNAVTL--ARNLRFSLGKYNFLFDDL 425
Db 105 NNNDNNNNFNNYS-DEIS-----KNIIHKDNLENLQKDLTKSISLKNKIVNYESKIEEL 158
Qy 426 ASHL-----DYYFLVSKAKIKQSSITKKLFIELPIKISLKSISLQDQPNL--- 471
Db 159 EKELKEVKDKNIDNDY---ENKLKEKEDFVKQKIDM-----LNEKENLLQEKELDINKR 210
Qy 472 -KTLFEKE--VTFKLDNFRDVEIEKAFGLLYPGVNEEQAQKARAFSEKEKKGK-K- 527
Db 211 EKKINEKEKNIIEKETTFHNI-----KEYLEKNNKERETISIEIIDIKKHLEK 258
Qy 528 ---EFQOKEE---NSKAINNOEGLEE-----DDNITERLPENSPYOYQ 565
Db 259 LKIEIKKKEDLENLKNLKNLKENVLEKGCVKKEKNETINSNDNIIIEKKYKLLLEYE 318
Qy 566 QENAGLAGS-PDKPYMTKDVQNYRYLAKSQIOELIKAKOYTKLAKLLSNRHTNLSRL 624
Db 319 LEENKQIDILNKQKEKEKEKEK-----EKEKEREKEKEKYDTLKLKXDE-----KISI 370
Qy 625 KEQLFDVNPRIIPSRDIEKAFVLDKTEKNKYQIYSSAPVFNQKWSLFGYRYLLGLD 684
Db 371 LEKVHSIKVR---EMDIEKRE----- 388
Qy 685 PKQTIHELVKLGQKAGLQFEGYENLPSDFNLEDLKNIRIKTIPFSLQKDNFKSLDLPNNY 744

CC	specification but is based on sequence information supplied to Derwent by			
CC	the European Patent Office			
XX				
SQ	Sequence 1948 AA;			
	Query Match	3.0%;	Score 288;	DB 5; Length 1948;
	Best Local Similarity	19.3%;	Pred. No. 1.8e-07;	
	Matches 348;	Conservative 298;	Mismatches 676;	Indels 484; Gaps 90;
QY	9	LLATAAAIIGSTVFGTVVGLAGKYKRGVNPQTQGVISQGLIDSVAFKPSIAFTSD---	65	
DB	339	LLVEAFKMGFT-----TNKNF--IQVLAIIHLGNITFTSWKSEQANFTNDSP	388	
QY	66	-----YQSVKALLNGKTFDPKSESEFTDFVSKFEDFLTNGRTVLEIPKQYVVISSE	120	
DB	389	DRICELAVDKOLFVQNLRLPKVAGREFITK-----SKKPNVKKFAIDAFK-	436	
QY	121	DDKERFRGLFHKEKLEDGNIAGSATFIYLLPLDMPKAAALQGYSYIVDKNFNLIHPL	180	
DB	437	-----YLYEKLFOFIITKINEK-----LDQSHGDDTANNNDNFFIGVL	478	
QY	181	-----SNFSAQSIKPLATRSSDFIAKLNQFNQDELWVYLEKPDFLEALKANIRLOTAD	235	
DB	479	DIAGEIIFDINSPEQLCINYTNE--KLQOFFNHHSF--ILEQ--SEYLRNINWEFID	530	
QY	236	FSFEKGNLVD-----PFVYSFIRNPQNKQWASDLNQDQKTVRLYLRTFSPQAKTILK	289	
DB	531	FGQDLQPTIDLIETKQPMGILKLDBECLMPKSSDASFMFKLSKNFTNTH-----	580	
QY	290	DYKYKDETFLSSIDLKASNGTSLFANENDLKOQLD-----VDLLD-----VSDYF---	334	
DB	581	-KKFSENKFGNGFIHHYAGKVEYVNWLNQNTDPISESLNLLPDSQNELIADMFIN	639	
QY	335	-----GGQSETITSNSQVVPASERSLKRVPKFKDQOK-----PRIEK	374	
DB	640	PHINRPQTNG-----GNSKLK--TASQKH-KDQLKTLMDQLESTEPHFVRCILENLEK	690	
QY	375	FSLYEYDALSFYSQLOELVSPNSIKDLVNATLARNLRFSLGKY--NFLPDDLASHLDYY	432	
DB	691	-RANKFDKNLVGLQR-----CNGVL-EGIRITRAGVYNNRMFDEF---IQRY	733	
QY	433	FLVSKAKIKOSSITKFLPIELPIKLSKSSILGDOEPNI---KTLFEKEVTFKLDNFRDV	489	
DB	734	SIICDNEL--SSPQNTNCE---IILPKVLNPEDFKVLTKIPFKNGILGKLEIRDL	787	
QY	490	EIEKAPGLLYPGVN-----EELQARKAQRASFEEKSKG---LKFEFSQOK-	533	
DB	788	ALKNIFTDLQKVIKGNLRLVLKQIKIEQSAQIISRTWVTLDEIKSNSPWRLLPFHVKP	847	
QY	534	--EENSKAIN--OEGLB-----EDDNITERL--PENSPYOYQOEN-----AGL	571	
DB	848	LLEDQSAVLDSKGLQENLQTLTVKLDSEKLTGKLETDNEKLREQMNLQDEMINITIA	907	
QY	572	GASPKPYMKDVQVQRYLAKSOJQL-IKAKDVTYKLAKLSNRHTYNISLRLEKQSLFD	630	
DB	908	KEKDKLSQLRLTENK---SKHRIETLEIKLDFEKQNDLINEHE---KLTRSLR	958	
QY	631	VNPRISSRDIEKAFVLDPKTEKNKYQIYSSASPVFQNKWSLFGYRYRLGLDLPQTIH	690	
DB	959	LH-----DKHNAKV---EELNSLNKLDHSAQSELSNL-----KQQLK	992	
QY	691	ELVKLGQKAGLOFEGYENLPSPFNLEDLK-----NIRIKTPIFSQKD-----	732	
DB	993	EITKLQNAHSKELLKPKQM-HDKSIEDLNKVSLENLNKKLQADLNSKQVVISSEHGKL	1051	
QY	733	-NFKLSLLDFNYYDGEIKAPE---FGLPFLPKELRRNSNGSGSONSPWEQEI---	785	
DB	1052	QNEIMKLELCSDYERKQDSQRTDLSLQIQLKKD-----EDYKSYEKIEEA	1100	
QY	786	-----ISQFKDQNLNQDLAQFSTKIWEKIIGDENEFQDQNNRLOYKL	828	
DB	1101	REKVTLKKGVEKKSQEIQYKSEIKKLAELASNTKI-----TQJDSQTK-ELSS	1151	

Qy	829 LKDQESW--INKTRDNLWYTLGDKLVKPKNNLEAKFR-----QISNLOELLTAFYTS	881
Db	1152 LKSNESKYVAIESAKQOL-----SKKIEYSSIEADYNRLQTELKQMKQNTVEYSTR	1204
Qy	882 -AALSNNWNYQDSGAKSTI-----IPFEIAELDPKVKKEKGVADYVQLKPHYAIG	930
Db	1205 ITELSTKLQVTEEAKEKIEKENQPPNPQFMEEPTHMKLVKNEQT-ASLRKEKFENK-K	1262
Qy	931 FPDNAGKFQVQEVIRSSRTIYKTS-----GSKLEADTIDOLNOAVKNAPLGLQSFYL	984
Db	1263 LSEELQMLKERINMGSLTSMDLTPKRRSLAIGDKSMITNTVDSFNKEIENLKQLO---	1318
Qy	985 DTERFGVQFKLATSIAVQHKQKTKLPKLNNDGYTLIHDKLKPKVIPQIPISSPEKDWFE	1044
Db	1319 --OEQGNFOR-AENYAI---ELQKLNKLTLTGLTNTD-----YE	1354
Qy	1045 GKLNQNGQSONVNVSTFGSIIESPYSFNFQEDADLDQDQD--DSRQ-----GNSLDN	1097
Db	1355 KKYNDQSKRITQLETIKIGRLAN---SGGDYEDHLLDDNDWESSRNSGSGGSGTSSS	1411
Qy	1098 QEAGLLKQKLAILLGN-----QFTQYQOQNDKEIEF--EIIINVEKVSLSFRV-----E	1144
Db	1412 ARNSLVKSESITAFHNMGRVSDYIQIYQDITTKLSTREELNGSKTEILRLKALLRESE	1471
Qy	1145 FKLAKTLEDNGKT-IRVLSDETMSLIV--NTTIKTPMSAVPEVDFTKWQVQDPRTPPL	1201
Db	1472 DELYQVQKQNYKTSVHDYEQDLAQLKVCHETLSRNKDINESLEIY-KRSDEYKKLEL	1530
Qy	1202 AAKTKFVLK-FKDQIPVDGSGNISDKWLASIPLVIHQOMLRLSPVVVKTIRLGLKTEQOQ	1260
Db	1531 ABSAIAISKRHEEQATKEMKESRSQ-----LLLVREELRTTQIL--IKDFRIKVENLE	1581
Qy	1261 QQQQQQQQQQPOKAVRKEELETYNPKDEFNILLN---PLTKAHLRLTSLNLVNDPNYK	1316
Db	1582 ATIEEKHQLDANK-----EETIKQIQDKLNLHKNFENKELNEKLEKEIKNL-NRDLDFK	1635
Qy	1317 IE-DLKVIKNEAGDHQLAFS---LRANNI-----KRLMNTPTTFADYNPPFYNN	1361
Db	1636 TDITETKLKIK-ENKKQLQDYEDVLLVKNLQNEVEELILOBEKLN-----	1679
Qy	1362 EDWRSIDKYLNNKNGVSSHQOQAGGQSGGLIQLRNKIKPKETFTPALIALKDRNNTNL	1421
Db	1680 ----KIDELTNNRNQLEN-----EKLINERKIVNCTKQISGLKE-----	1714
Qy	1422 SNYSDKIIMIKPKYLVERSIGVPWSTGLDGVISGEQTK-DGTSSSSQKGFDDQFIO---	1477
Db	1715 --LVDEISIERDKLLADKET-----LQNDLQTMKTFADATTTELKQAHGELDFLKXHL	1765
Qy	1478 -----ALGLKQNTVEYHGKGLGLSIRIFDP--GNELAKIKDAS---NKGGEKLLKSYDL-	1524
Db	1766 ENQREDSEAIKTELNQSKMSTSFDIRDQOQKLRNELLVTKBEENFSLVKTNKLNLKVSdle	1825
Qy	1525 PKNYLNE---YEKKSPKIAKGTWNIIHPDQKEYPNPQKLPENYLNL---VLNQPKWYTL	1577
Db	1826 EKLYSNEQLKYWESKVDTLSKALDGLALNEKHEADTKIKLQRSIKOLEIRVENESQLSKR	1885
Qy	1578 YNSDDF 1583	
Db	1886 YNDENF 1891	
RESULT 22		
ADB75595		
ID	ADB75595 standard; protein; 1979 AA.	
XX		
AC	ADB75595;	
XX		
DT	04-DEC-2003 (first entry)	
XX		
DE	Prostate cancer marker protein.	
XX		
KW	Prostate; cancer; cytostatic; gene therapy; marker.	
XX		

OS Homo sapiens.
XX WO2003009814-A2.
XX 06-FEB-2003.
XX 25-JUL-2002; 2002WO-US023913.
XX 25-JUL-2001; 2001US-0307982P.
XX 25-AUG-2001; 2001US-0314356P.
XX 25-SEP-2001; 2001US-0325020P.
XX 12-DEC-2001; 2001US-0341746P.
XX 05-MAR-2002; 2002US-0362158P.
XX (MILL-) MILLENNIUM PHARM INC.
XX Schlegel R, Monahan JB, Endege WO, Gannavarapu M, Gorbacheva B;
PI Hoersh S, Kamatkar S, Wonesey AM, Glatk K, Zhao X, Anderson D;
XX WPI; 2003-248033/24.
XX New nucleic acid molecule, useful for diagnosing or treating prostate
PT cancer.
XX Disclosure; SEQ ID NO 419; 99pp; English.
XX The invention relates to newly discovered cancer markers associated with
CC the cancerous state of prostate cells. Also disclosed is a method of
CC assessing whether a patient is afflicted with prostate cancer. The method
CC of the invention involves assessing whether a patient is afflicted with
CC prostate cancer by comparing the level of expression of a marker in a
CC patient sample and the normal level of expression of the marker in a
CC control non-prostate cancer sample, where a significant increase in the
CC level of expression of the marker in the patient sample and the normal
CC level indicates that the patient is afflicted with prostate cancer.
CC Nucleic acids of the invention are useful for diagnosing or treating
CC prostate cancer, and may be useful in gene therapy. Sequences given in
CC ADB75177-ADB75631 represent marker cDNA and proteins. Note: The sequence
CC data for this patent did not form part of the printed specification, but
CC was obtained in electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences.
XX Sequence 1979 AA;
SQ
Query Match 2.9%; Score 286.5; DB 7; Length 1979;
Best Local Similarity 19.0%; Pred. No. 2.2e-07;
Matches 373; Conservative 345; Mismatches 682; Indels 563; Gaps 99;
22 FGTVGLASKVKYRGVNPOTGVISQLGLIDSVAFKPSIANFTSD-----YQSVKKAL 73
164 FGDIISSQQEIN-RLSNEVSRLESEVGHWRHIA-OTSKAQGTDSQSEICKLQNIKEL 221
74 LNKGTDPKPSSEFTDFVSKPFDLTN-NGRTVLBPKKYQVVISSEFSPEDDKERFLGPHL 132
222 KQN-----RSQEIDDDQHESVVLQNAHQKQLTISRHRREELSDY-----EERI----- 265
133 KEKLEPGNTAQSATKIYLLPLDMPKAAALQGVSYIVDKNFNNLIHPLNFSQAQSIKPLA 192
266 -ELE--NLLQOGSGV--IETDLSKI-----YEMQKTIQVLQIEKVE--STPKMEQLE 312
193 LTRSSDFIAKLQFNPN-----QDELWYLEKFPDLEALKANIBLQ-----TA 234
313 -DKIKDINKKLSAENDRDILRREQSQLNV--EKROIIMECE--NLKLECSKLOPSAVKOS 368
235 DFSFEKGNLV-----DPP-VYSFIRNPONQEWASDLNQD-----QKTVRLYLRTEFSP 282
369 DTWTEKERILAQASVEEVRFLQALSDAENEIMRLSSLNQDUNS LAEDNLKMKRIEVL 428
283 QAKTILKDYKYDETFLSSIDLK-----ASNQTSIPANENDLKQDLDVLDVSDY 333
429 KEKSLLS--QEKEQLMSLLKLNNEYEVIKSTATRIDSLDSELHLRLNLE----- 477
334 FGQSQSETITSNSQVQVPVPSASERSLKDRVKFKDQKQPKRIEKFSLYEYDALSFYSQLELV 393

Db 478 -----AKEQLNQSGIS-EKETLJAEIEELDRQNEQATKMILIKQOL 518
QY 394 SKPNSIKOLVNATLARNLRFSLGKYNFLPD---DLASHLDYVFLVSKAKIKOSSITKFLF 450
Db 519 SKQONEGDSIIISKLKQDLNDEKKRVHOLEDDKWDITKELD---VQKEKLIQSEVA--- 570
QY 451 IELPIKISUKSSILGQEPNITKLPKEVTVFKLDNPRDVEIEKAFGLLYPGVNEELEEQAR 510
Db 571 -----LNDLH---LTKQKLEDKVENL-----VDQLN 593
QY 511 KAQRASFEKEKSKGLKEFSQOKEENSXAINNQ--EGLEEDN-----ITER----- 555
Db 594 KSOENSVSIOKENLELKEHIRQNEBELSRIRNELAQSLNODNSNFKDTLLKEREAEVRN 653
QY 556 LPEN-SPIOYQOENAGLGLASDPKPYMIKDVQORVYLA---KSOIQELIKAKDYTKLAK 610
Db 654 LKQNLSELEQLNENLKVAPDVK-----MENEKLVLACEDVRHOLEECLAGNQLSLEK 707
QY 611 LLSNRHTYINISURLKEQLFDVNPRIIPSSRDI EKAKFVLDKTEKNKYQIYSSASPVFQNK 670
Db 708 -----NTIVETLMEKEGEIEAELCWAKKRLLEA-----NKYEKTEIELSNA-RN- 751
QY 671 WSLFGYRYVLLGLDPKQTIHE-LVKLGOKAGLOF---EGYENLPSDF-----NLE 716
Db 752 -----LNTSALQLEHEHLIKLNQKQMEIAELKKNIEQOMDTDHKTKDVLSSSLE 801
QY 717 DLKNIRIKTLPFSQKDNFKLSLLDFNNYVDGIGIKAFEGFLPLFKELRRN----- 767
Db 802 EQQL---TOLINKKEIFEIKLKERSSKLOELDK-----YSQALRKNELRQTIEE 850
QY 768 SNSGGSQNSNFWQEIISQFKDQ-----NLSNQOLA-----QFST----- 805
Db 851 KDRSLGSMKEENHHLQELERLEREQSRTPAVADPKTLDSTVTELASSEVSLNTIKEHLEE 910
QY 806 --KIWEKIGDNEFDQNNRLQYKLLKDIQSSWINKTRDNLVYTLGDKLVKPKNNLEA 863
Db 911 EIKHHQKIIEDNQ---SKMQ--LLQSLQEO--KKEMDE--FRYQHEQMNATHQTLFLE 960
QY 864 KFRQISNLOELLTAFYT-----SAALSNNMNYQDSGAKSTI 900
Db 961 KDEEIKSLQKTIQIKTQLHEERQDIQTDNSGDI FQETKVQSLNIENGSEKHDLSKAETER 1020
QY 901 IFEETAELDPKVK---EK---VGADVYQLKHYAIGFDDNACKFQEV----- 942
Db 1021 LVKGIKERELEIKLNEKNISLTKQIDQLS-----KDEVGKLTQIIQQKOLEIQALHA 1073
QY 943 -IRSSSRT---LYLKTSGSKLEADTIQO-----LNOQAVKNAPLGLQSYLDTFRFVGF 992
Db 1074 RISSSTHTQDVVYL---QQQLQAVAMEREKVFVAVLNKTR-----ENSHLKTVEHKOM 1123
QY 993 QKLATSLA--VQHKQKERTLPKKLNNDGYTLIHDKLKPKVPI PQISSSPKDFEGKLNQ 1050
Db 1124 DIVAAKEAALIKLDENKSLSTRFESSGQDMFRETION--LSRIIR--EKDIEDALSQK 1179
QY 1051 QGS-----QNVNVSTFGSIIIESPYSTNFQEDADLD-----QDGQDS 1088
Db 1180 CQTLAVLQTSSTGNEAGVNSHQFEELOERDLKQVKKMEWKKQVMTTVOQNMOHES 1239
QY 1089 RGNNSLDNOEAGL---KOKLAILLGNOPFYVQQNDKEI----- 1126
Db 1240 AOLQELHQLQAOVLVSDNNSKLQVDY-TGLIQSYEQNETKLKNPGQELAOVHISIGL 1298
QY 1127 -----EFBIIN-----VEKVSLSFRVEFKLAKTLSD 1153
Db 1299 CNTKDLLLGLDIIISPOLSSASLLTPQSAECLURASKSEVLSSELLQOELBELRKSLOE 1358
QY 1154 NGKTTIRVLSDE---TMSLIVNTTIEKTPENSASVEPDTKWEQVDPRTPLAAKTFVL 1209
Db 1359 KDATIRTOENNRHLSDSIATSELERKEHEQTDSEIKQLK--EKODVLQKLKEDLLI 1416
QY 1210 KFKDQIPVDGSGNISDKW-----LASIPLVIHQMLRL-----SPVVKTIRELG 1253

Db 1417 KAKDQLSSNENFTKVNENELLRAQVATNLKERILILEMDIGKLGKGENKIVETYR -G 1474

Qy 1254 LKTEQOQOQOQOQOQOQKAVKKEBELETYNPKDEFNINLPJTKAHLRLTSLNVLNNDP 1313

Db 1475 KETEQALQETNMKFSM-----MLREKEFECHSMKALAPEQLLKEQKGTGELNQLL 1529

Qy 1314 N--KYIEDLKVYKNEAGDHQALAFSLRANNIKRLMNTPTTFADYNFFPFYNNEDRSIDKYL 1371

Db 1530 NAVSMQSKTVVFOQERD-QVMLALQKQ-----KOMENTAL-----QNEVQRLDK-- 1573

Qy 1372 NNGKNVSHQOQAAGNGGSLIORLNKNI--KPEFTFPALIALKDRNTNLSNYSKII 1429

Db 1574 ----EFSNQE-----LERLRNHLESSEDSYTREALAAEDRE----AKLRKKVT 1614

Qy 1430 MIKPKYLVRSIGVPMSTGLDYGSGEQTKDGTSSSQOQKGFDDQFIQ-----A 1478

Db 1615 VLEER-LVSSSSNAME-NASHQASQVESLOBQLNVSKQR--DETALQLSVSQBKQYA 1670

Qy 1479 LGLKNTYHKGKLSIRIFDP-----GNEIAKID--ASNKKGEELKLSYDLFKNYLN 1530

Db 1671 LSLAN-----LQWLEHFOQEEKAMYSAELEKQKQLIAEWKKAENLEGKVISLQECID 1724

Qy 1531 EYEKSPKIAK--GWTNIHPDQKYPNPQKLPENYLNVLNQPMKVTLYNSSD----- 1582

Db 1725 EANAALDSASRLTEQLDVKEQIEELKEQNELRQEMLDV--QKKLMSLANSSEKVDKV 1782

Qy 1583 FITNLFV-----EPEGSDR-----GSGTKLKQVIOKQVNNVADWGS--VLTFWY----- 1626

Db 1783 LMRNLFIGHFTPKNQREVLRLMGSILGVRREMEQLFHD--DQGSVTRMTWGLGGGS 1840

Qy 1627 ----DKNITQPNVITANIADVFKDVKLEEDNTKLIAPNIT 1665

Db 1841 KSPVNTLRPNQOQSVVNSFSSELFVKFL-ETESHPSPIPPPLIS 1882

Query Match 2.9%; Score 281; DB 6; Length 1805;
Best Local Similarity 19.2%; Pred. No. 4.1e-07;
Matches 361; Conservative 282; Mismatches 627; Indels 614; Gaps 93;

Qy 77 KTFDPKSS-----EFTDFVSKFDFLTNNQRTVLEIPKPKYQVVISFSPEDDKERPLG 129

Db 2 KPFDKPSLQPIYDIGDFDDGYLQSEYKNSKT--DVDKIENQLLKEIKSLEDELNKLKG 59

Qy 130 FHAKELKEDGNIAQSAATKFIYLLPLDMPKALQGVSYVDKNFNLL-----IHLPLSNF 183

Db 60 --LKNQAEIN-----PE-----LDKKINHLEVDLNRNLVNEYKNF 91

Qy 184 SAQSIKPLALTRSSDFIAKLNQFNQDELWYLYLEKFFDLEALKANIRLQTDADFSEK-GN 242

Db 92 QFQ-----KNHWDVKVSELDN-----LTFYKNELTR--LQENADFLNSKYAN 133

Qy 243 LVD-----PFVYSFIRNPQKQEWASDLNQ-----DQKTVRLYLRTESPOAKTILKDYK 292

Db 134 LANFOANYHNKLNDFHRLIENQNTINRLNKGNGQNQLIDNNVALLQNPITVEKKNY- 192

Qy 293 YKDETFLLSSIDLKASNGTSLPANENDLKQDLVD--LLDVSDYFGGQSGSTTSNSQVXPV 350

Db 193 -----LNVID-----OLY-----NEL--DQLEENOKRLSI--EYENTYRELVSADNELQNV 235

Qy 351 PASERSLNDVRVKFKDQKQKPRIEKFSLYEYDALSPYSOLOELVSKPNSIK-----DLV 403

Db 236 --YENIDQNOIQFKHOYQTYRELSQLERKILT-----KQELVDKESALRVKIDDAFVI 289

Qy 404 NATLARNLRFSLGKYNFLFDDLASHLDYVFLVSKAKIKQSSITKFLFELPIKLSKSSI 463

Db 290 NARLAE-----LDDVAKQLSF-----QDGIITKQNAQHVDEKLL-----VA 323

Qy 464 LGDQEPNITLFEKEVTFKLNDFRQVETEKAFGLLYPGVNEELEQAKAQAASPEKSKS 523

Db 324 LNKEKDRLNT--QKEAFFNLRSQALIDLNK-----LQOENELFAKHLHQHQQNEFFQKQSD 376

Qy 524 KGLK-----EFSQKKEENSKAINNQE--GLEEDDNITERLPENSPIOYQENA 569

Db 1417 KAKDQLSSNENFTKVNENELLRAQVATNLKERILILEMDIGKLGKGENKIVETYR -G 1474

Qy 1254 LKTEQOQOQOQOQOQOQKAVKKEBELETYNPKDEFNINLPJTKAHLRLTSLNVLNNDP 1313

Db 1475 KETEQALQETNMKFSM-----MLREKEFECHSMKALAPEQLLKEQKGTGELNQLL 1529

Qy 1314 N--KYIEDLKVYKNEAGDHQALAFSLRANNIKRLMNTPTTFADYNFFPFYNNEDRSIDKYL 1371

Db 1530 NAVSMQSKTVVFOQERD-QVMLALQKQ-----KOMENTAL-----QNEVQRLDK-- 1573

Qy 1372 NNGKNVSHQOQAAGNGGSLIORLNKNI--KPEFTFPALIALKDRNTNLSNYSKII 1429

Db 1574 ----EFSNQE-----LERLRNHLESSEDSYTREALAAEDRE----AKLRKKVT 1614

Qy 1430 MIKPKYLVRSIGVPMSTGLDYGSGEQTKDGTSSSQOQKGFDDQFIQ-----A 1478

Db 1615 VLEER-LVSSSSNAME-NASHQASQVESLOBQLNVSKQR--DETALQLSVSQBKQYA 1670

Qy 1479 LGLKNTYHKGKLSIRIFDP-----GNEIAKID--ASNKKGEELKLSYDLFKNYLN 1530

Db 1671 LSLAN-----LQWLEHFOQEEKAMYSAELEKQKQLIAEWKKAENLEGKVISLQECID 1724

Qy 1531 EYEKSPKIAK--GWTNIHPDQKYPNPQKLPENYLNVLNQPMKVTLYNSSD----- 1582

Db 1725 EANAALDSASRLTEQLDVKEQIEELKEQNELRQEMLDV--QKKLMSLANSSEKVDKV 1782

Qy 1583 FITNLFV-----EPEGSDR-----GSGTKLKQVIOKQVNNVADWGS--VLTFWY----- 1626

Db 1783 LMRNLFIGHFTPKNQREVLRLMGSILGVRREMEQLFHD--DQGSVTRMTWGLGGGS 1840

Qy 1627 ----DKNITQPNVITANIADVFKDVKLEEDNTKLIAPNIT 1665

Db 1841 KSPVNTLRPNQOQSVVNSFSSELFVKFL-ETESHPSPIPPPLIS 1882

RESULT 23

ABU35589

ID ABU35589 standard; protein; 1805 AA.

XX AC ABU35589;

XX 19-JUN-2003 (first entry)

XX Protein encoded by Prokaryotic essential gene #21116.

DE Antisense; prokaryotic essential gene; cell proliferation; drug design.

KW Mycoplasma genitalium.

OS WO200277183-A2.

XX 03-OCT-2002.

XX 21-MAR-2002; 2002WO-US009107.

XX 21-MAR-2001; 2001US-00815242.

PR 06-SEP-2001; 2001US-00948993.

PR 25-OCT-2001; 2001US-0342923P.

PR 08-FEB-2002; 2002US-00072851.

PR 06-MAR-2002; 2002US-0362699P.

XX (ELIT-) ELITRA PHARM INC.

PA Wang L, Zamudio C, Malone C, Hasebeck R, Ohlsen K, Zyskind JW;

PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;

XX WPI; 2003-029926/02.

DR N-PSDB; ACA39459.

XX New antisense nucleic acids, useful for identifying proteins or screening

PT for homologous nucleic acids required for cellular proliferation to

PT isolate candidate molecules for rational drug discovery programs.

XX

Query Match		2.9%, Score 281, DB 7, Length 1805;
Best Local Similarity		19.2%, Pred. No. 4.1e-07;
Matches		361; Conservative 282; Mismatches 627; Indels 614; Gaps 93;
Qy	77	KTPDPKSS-----EFTDFVSKFELTNGRTVLEIPKYYQVVISFSPEDDKERPLG 129
Db	2	KPFDPKPSLOPIYDIGDDGVLOSEYKRNKST--DVDKIENQLLKKEIKSELEBELKMLKG 59
Qy	130	FHLKEKLEDGNGIAQSATKFIYLLPLDMPKAAALGOYSYIVDKNFNL-----IIHPLSNF 183
Db	60	--LNQAEDN-----PE-----LDKKINHLVDLNRNLVNEYKNF 91
Qy	184	SAQIKPLALTRSSDFTAKLNQFNQDELWYLEKFFDEALKANIQLTADSFKEK-GN 242
Db	92	QFQ-----KHMVDVKSELND-----LTFYKNELTR--LOQENADFLNSKYAN 133
Qy	243	LVD-----PFVYSFIRNPONKEMASDLNQ-----DOKTVRLYLRTBFSQAOKTILKDYK 292
Db	134	LNFQANVHKLNDPHRLIENQNTINRLNOKINGNQLNDNNVALLQNFNITVEKQNY- 192
Qy	293	YKDETFLSSIDLKASNGTSLFANENDLKQDLVD--LLDVSDYFGGSETITSNSQVKPV 350
Db	193	-----LLNVID-----QLY--NEL-DQLENQKRLLSI-EYENTYRELVSADNELQNV 235
Qy	351	PASBSRLKDRVKFKDQOKPRIKPSLYEYDALSFYSQQLVLSKPSIK-----DLV 403
Db	236	--YENIDQNIQFKHQYQTYRDELSQLERKTLT-----KQELVDKESALRVKIDDAFYI 289
Qy	404	NATLARNLPISLKGYNFLFDLASHLDYVFLVSKAKIKQSSITKKLFTLPKISLSSSI 463
Db	290	NARLAE-----LDDVAKQLSP-----QDGIYQNAQHVEDKL-----VA 323
Qy	464	LGDQEPNITLFEKEVTPLKDNFRDVEIEKAFGLLYPGVNEBELQARQAASPEKSKS 523
Db	324	LNKEKORLNT--QKEAFNLQASALIDINK-----LQENELFAKHLHQNEFEQKQSD 376
Qy	524	KGLK-----BFSQKEENSKAINNOB-GLEEDNITERLPENSPIYOQENA 569
Db	377	SLLKLEYKALQHKINEFKNESATKSEBELLNQRELFEKREIDTLLTQASLEYHQ- 435
Qy	570	GLGASDPKPYMKDVQORYYLAQSIOELKA-KDVTKLAKLLSNRHTYNISLRLKEQL 628
Db	436	-----ESSQLLKQKNE-----VKQHFONLEVAKKELDKERNLLDQKKVD-----SEAI 480
Qy	629	FDVNPRIPSSR-DIEKAKFVLDTKEKNKYQIYSSASPV-----FQNKNSLFGYYRYLL 681
Db	481	FOLKEKVAQERKELEBYLVKKQKQDKENELLPEKQKQHQADPFENE----- 529
Qy	682	GLDPKQTIHELKQKAGLOPEGYENLPSPF-NLEDLKNIRIKTPLPSOKD-NFKLSLL 739
Db	530	-LEAKQ--QEL-----FEAKHALERSFIKLED-----KEXDLNTKAQOI 565
Qy	740	DFNNYDGEIKAPERGLPLFLPKELRRNSSNGSGSONSPWEQIIISQFKDQNLNSQDQ 799
Db	566	-----ANEF-----SQLTKDKSKA-----DFELMLQNEYENLQEQK 598
Qy	800	LAQFSTKIWEKIGDENEFQDN-----NRLQYKLLKDLQESWINKTRNLYWYTLGDKLK 854
Db	599	-----KLQERTYFERNAAVLNRLQKREELLQ-----KETLDLTKSPQERL- 644
Qy	855	VKPKONLEAK--FROISNLQELLTFATYSALSNWNYQDSGAKSTIIFEBIADLPKV 912
Db	645	-----INQREHKELVASVEKQKIL-----GKXLQDPSQTSLSNASKNLAEREMAI 689
Qy	913	KEKVGADVQLKFAHFGDDNAGKFNQBSVIRSSSRTTYLTKSGSKLEADTIDQLNOA- 971
Db	690	-----KPEKEIBATEKQLLDVNNAEVIOAD-LAQNLQSL 724
Qy	972	-----VKNAPLGLOSFYLDTRFVGFKL-ATSLAVQHKOKE-KTLPKKNLNDGYTIIH 1023
Db	725	NOERSELQNAQRIADFNHDS-----LKKLNEVELSLQKRLQELQTLLEANQKQHSY---- 775
Qy	1024	DKLKKPVIPOISSPEKDMFEG---KLNQNGOS-----QNVNVSTFGSIIESPYFSTN 1073
Db	776	-----NQAYFEGELDKLNEKQAFNLNRKKQTMEDVDAIKQRLSDKHQALN 821
Qy	1074	FQSDADLDQGDSDSGNNSLDNOBAG--LLKQKLAILLGNQFI----- 1116
Db	822	MQQ-AELDR-----KTHELNNAFLNHADQKSLQDQLATVKETQKLDLERSALLERQREF 876
Qy	1117	-----OYYQNDKEIFEI--NVEK-----VSELSFRVE 1144
Db	877	AENVAGFKRHWSNKTLOLOKIYELTKKQSEQTKETELKIAFSDQKQYQVFELOKQDE 936
Qy	1145	FK-----LAKTLSDNGKTIIRVLSDETMSLVN---TTIEKTPEMSAVPEVFDTKWVE 1193
Db	937	FROIKAQRELDKLAENNQ-VKLELDNRFQALNQKQDTVQAQLELER-----E 985
Qy	1194	QYD--PRTPLAAKTFVLKFDQIIPVDGSGNISDKWLA-SIPLVTHQMLRLSPVVKTI 1249
Db	986	QHQLNLEQTAFNAQANESLLKQREQ-----LTKKIQAFHYELKKNQFQALKG----- 1032
Qy	1250	RELGLKTEQOQOQOQOQ-----QOQOQOQKCAVRKE 1279
Db	1033	KRLFAKEQDQQRKQDQBNRPFQFEKEYTDFDAKRELEBELEKIRRSLSQSNVELEKRR 1092
Qy	1280	EELETYNPKDEFNINPL---TKAHLTLNLV-----NNDPNYK----- 1316
Db	1093	EKLAT-----DFTLNKKVQHTQINRDQLNSQIRQFLLERKNFQFSNEANAKKAPLKR 1147
Qy	1317	----IEDLVKINEAGDHLAFSLRANNIKR-LMNTPIITFADYNPPFYNNEDWRSIDKYL 1371
Db	1148	LRSFASNLIKQKALATQKLEFDRDEQKQKELQOATLQLEQF-----KFE 1193
Qy	1372	NNKGVSSHQOQAGNGSGSLIQRNLNKNIKPFTFPALIALKDRNNTNLSNYSDKIIMI 1431
Db	1194	KQNFIDIEKQRLVAIKTQ-----CEKLSDEK--ALNOKLVELKNLSQTYLANKN----- 1241
Qy	1432	KPKYLVERSIGVPMWSTGLDGYIGSEQTKDGTSSSSQKQDQDQIQAALGKNTYHOKLG 1491
Db	1242	KAEY-SQOQLOQKYTNLLDLKENLERTKD-----QLDKKHSIFARLTKFANDLRFKKQ 1295
Qy	1492	L--SIRIFDPGNELAKTKDAS-----NKGEEKLLKSYDL 1524
Db	1296	LLKAQRIVDDKNRLKENERNLHFLSNETERKRAVLEDOQISYEFKQKQATDAIASHKE 1355
Qy	1525	FKVLYNEYKSKPIAKGWTNIHPDOKEYENPNOKLPENVLNVLNPOWVTLNYSDDFI 1584
Db	1356	VKKEGELQKLLVELETRTKLNNDFAKRSQREEFENQRLKLELOKTLQOTQNSNNFK 1415
Qy	1585	TNLFVEPEGSDRSGTKLKQVIQKQVNNNYADMGSAYLTFWYDKNIITNQPNVITANIAD 1644
Db	1416	TKAQIETIENSVKRGMEELN--FQKKE-----FDKN-----KSRLYE 1449
Qy	1645	VFIK---DVKELENTKLIAPNITQWPNI---SGSKEKFKYKPTVFGNWNENSSMNSQ 1698
Db	1450	YFRMRDEIERKESQVKVLKE-TORKANILLEAQANKLIEKNTIDPFKEB----- 1499
Qy	1699	AQPTWEKIRGEFALQALKSSFDQ 1722
Db	1500	-----LKAFKQKVDQ 1509
RESULT 25		
Id	AA18161	standard; protein; 2441 AA.
xx	AA18161;	
xx	AC	
xx	DT	07-NOV-2000 (first entry)
xx	XX	Plasmodium falciparum chromosome 2 related protein SEQ ID NO:18.
xx	DE	Plasmodium falciparum; chromosome 2; human malaria parasite; vaccine;
xx	KW	antimalarial; malaria; protozoacide; infection; insecticide.
xx	KW	


```
Qy 602 -AKDYTKLAKLNRHTYNTSLRKE-----QLFDVNPRIPSRDIEKAKFVL 648
Db 927 LKXDY-QLPELQKQNFQIQEQKHQLELLAQOAEKQELKQATALASQDQDTQAKL 985
Qy 649 DKTEKNYWIYSSAPVFNKWSL-FGYRYLLGLDPKQTIHELKVLGQAGL----- 701
Db 986 DLARQQHELELRNWF-----NQASLSLNKQREQLTNQVKVYLHGELKKRHEKLTLDORLLA 1041
Qy 702 -----OPEGYENLPSDFN-----LEDLKNIRIKTLPFSQKDNFKLSLID 740
Db 1042 EKEKDQHKDAEINQRPQENYADPDQAKKELQELNQR-----RNLEOSNASLLK 1095
Qy 741 FNNYDGEIKAPFGPLFLPKELRRNSNGSSONSPEQEIISQFKD-QNLNQ-- 797
Db 1096 KRN-----QLTLDFAALLRVQHTNTQTRVQLNTQI---KEFLLEKKNFGKASDEAA 1143
Qy 798 -----DQLAQFSTKIW---EKIIGDENEFQNNRLOYKLLKDLQESWINKTRDNLWT 847
Db 1144 LQKALLIKLURSFAKLQREALATQKLEFDRD-----EQKSEINNAKLQ--- 1192
Qy 848 YLGDKLVKPKNNLEAKFRQISNLQELLTAFYTTSAALSNNWNYQDSCAKSTIIFEEIAE 907
Db 1193 ---EQFKLEKQNDQAKQKQLEIFKQOCQRLDVEKLLKQKLVQKNLSKSYLYTKNRAD 1249
Qy 908 LDPKVKKEGVADYVQLKFHYAIGFDNNAGKFNQEVIRSSRTIYLYKTSKGSKLEADTIDQ 967
Db 1250 LSQQQLQHVKYANLELK-----EKLQAKRALDKKHRAIYGMQAFVSE 1293
Qy 968 LQAVKNAPIGLQSFYLDTER-FGVQKLAISLAVQHKQKEKTLPKLNDGVTLLHDKL 1026
Db 1294 LROEKKQL-LSAQKQVDDKSRLEQQRHLQNLSSSTKKRQSLHDIK-----FDOR 1346
Qy 1027 KKPVIPOISSPEKQFEGKLNQ-NGOSQNV--NVSTFGSIIESPFSFTNFQDADLDD 1083
Db 1347 RKAHSVLSNKH-----KLKQEGELOILOKLSLKTQIEQE-FSKLYQOREKLDQ 1399
Qy 1084 QDDSRQGNNSLNOBAGLKQKAILLGNQFOIYYQOQNDKEIEFEIINVERKVSLSFRV 1143
Db 1400 RTTSLKHLRELKQAEATAHKNREVLEIEN-----YVKELQRLTTEKSEFDDNNKRLFEY 1455
Qy 1144 EFKLAKTLEDNGKTIKRVLSDETSMLIWTIETKPEMSAVPEVFDTKWVEQYDPRPLAA 1203
Db 1456 FRIRNEIEKKEAHIKTVLEETQK-----KRHLVETEAVKHLQKQSIIS 1500
Qy 1204 KTKFVLKFKDQIPVDSGNSISDKWLASIPLVIHQOMRLSPVVKTIKRELGLKTEQOQQQQ 1263
Db 1501 KGOELKEIKERSVSDISH--TNKQRELSLHQLKQNLKLAEREREINNKSLSLTQKI 1558
Qy 1264 QOQOQOQOQKKA-----VRKBE-----LETYN-----PKDEFNINPLTKAHL 1303
Db 1559 QTAQKLSKEARILKLEKRAVEQOQYQAEITRLKTRNADLEKDNKHLFPPLPKIN-- 1616
Qy 1304 TLSNLVNNDPNY-----KIEDLVKNEAGDHQLAFSLRANNTKRLMNTPTIFA 1352
Db 1617 -----GNDMNYPYFPWFYPOQKQSD-----SSNQIRHLFEQQLQFM 1653
Qy 1353 DYNPFYFYN--DWRSDTKYLNK-GNVSSHQOQAAGNQGSLIQLRLNKNI-----KPE 1404
Db 1654 QQR---YENELTELQRORALLEKLLQIQLESQLSAKKNDPEKVEQMQLLEKTEQKLS 1710
Qy 1405 TFTPALIALKORNTNLSNYSVD 1426
Db 1711 AFDQKINALAEQINTQKAEHAD 1732

RESULT 27
AAW20828
ID AAW20828 standard; protein; 2440 AA.
XX
AC AAW20828;
XX
DT 16-JUL-1997 (first entry)
```

```
XX
DE H. pylori cytoplasmic protein Ilgel0309orf15.
KW Cytoplasmic; vaccine; prevention; treatment; infection; identification;
KW binding compound; bacterium; life cycle; activator; bacteria; inhibitor;
KW duodenal ulcer disease; chronic gastritis; diagnosis; envelope.
XX
OS Helicobacter pylori.
XX WO9640893-A1.
XX 19-DEC-1996.
XX
XX 06-JUN-1996; 96WO-US009122.
XX
XX 07-JUN-1995; 95US-00487032.
XX
XX 01-APR-1996; 96US-00630405.
XX
XX (ASTR ) ASTRA AB.
XX
XX Smith D, Berglindh OT, Mellgaard BL;
XX
XX WPI: 1997-052306/05.
XX
XX N-PSDB; AAT68081.
XX
XX Helicobacter pylori nucleic acid sequences and related polypeptide(s) -
XX useful for vaccines to treat or prevent H. pylori infection, and to
XX detect Helicobacter.
XX
XX Claim 61; Page 1233-1238; 1481pp; English.
XX
XX The present sequence shows a Helicobacter pylori cytoplasmic protein that
XX may be used in a vaccine to prevent or treat H. pylori infection or to
XX identify H. pylori polypeptide binding compounds, useful as potential H.
XX pylori life cycle activators or inhibitors. The genomic sequence of H.
XX pylori (ATCC 55679) was determined from overlapping contigs generated by
XX mechanically shearing the bacterial DNA. The sequences were analysed for
XX ORF of at least 180 nucleotides, and the predicted coding regions defined
XX by computer evaluation. To identify likely H. pylori antigens for vaccine
XX development, the amino acid sequences predicted from various ORF were
XX analysed for significant homology to other known or exported membrane
XX proteins. Having identified and determined the sequences of interest,
XX particular regions can be isolated from H. pylori by PCR amplification
XX for recombinant polypeptide production, e.g. in E. coli hosts
XX
XX Sequence 2440 AA;
```

```
Query Match 2.9%; Score 280.5; DB 2; Length 2440;
Best Local Similarity 18.5%; Pred. No. 6,9e-07;
Matches 385; Conservative 316; Mismatches 678; Indels 697; Gaps 100;

Qy 41 QGVISQLGLID----SVAFKPSI--ANFTSDYQSVKALLNGKTFDPKSEFTDFVSKFD 94
Db 593 QG-LQQLGFNNDNHPKIFEPSLGTGKFTAHAPSCKNYRFIGTLPDIPISANLSKFLYPNQ 651
Qy 95 FUTNNGRTVLETPKXYQVVIS-----EF9BEDDKE-----RPLGLHLEKLEDDG- 139
Db 652 VIQNTALENYQFYQBYDAFVGNPPYGNHKKIYSSNDKELSNESIHNYFLGKAIKELKDDGI 711
Qy 140 -----NIAQSATKFIYLLPLDMPKA----- 159
Db 712 GAFVSSWFMMDAKNPMREHIAKNAT---FLGAILPNVSVFKATGAEVTSDIVFPKKGVE 768
Qy 160 -----ALGOYSYIVDKNFNNL---IIHPLSNFSAQSIKPLALTRSSDFIAKLNQFN 207
Db 769 KATNQSFTKAMPY---DKILNSLDDDTLFAIQNNRPDSFIP-----SDQLKIVNAVA 818
Qy 208 N-----QDELWYLEKFFDLEALKANIRLOTADFSF--EKGMLVDPFVYSFTRNPQOKE 260
Db 819 NHFGFKQEKLRWYEK-----IDTANFGYSTQDYKIIFDI-----DKV 857
Qy 261 WASDLNODQKTVRLYLRTFSPQAKTI-----LKDYKYKDETFLLSI---DLKASNGTSL 312
```

Db 858 GKSNINNEQTLNEY-----FIHPENILGHLSEKTYRRETNGEQIYKYDLOALEDESL 913
Qy 313 FANENDLKQDLVDLLDVSYFGQSETIITSNQVQVPVPAERSLKDRVRFKDKQKPRI 372
Db 914 DLSQ-ALKQAEIKLPKDVQYQY-----HKTTLKTDVLIIIDSNERYQEVQKLKN-----L 962
Qy 373 EKFSLEYDALSFYSLQOE-----LVSKP-----NSIKDLVATL- 407
Db 963 ERRELVKMDNL--YFQLEQNNEMGIFLKPTKINSKVQDSRLKAYFKIKDALNDLTSALN 1020
Qy 408 ----ARMLRFSIGKYNFLFDDLAGHLDYFVLSKAKIKQSSITKKLFIELPIKLSKSSI 463
Db 1021 PLSSDLELENKRAKLNVLVDFVKKFGY-----LNENKVK-----DIRQDL 1062
Qy 464 LGDQEPNIKTLEKEVTFKLDNFRDIEKAFGLLYPGVNEELEQAKAQAASPEKEKSK 523
Db 1063 YGAKVLGLEKDFEKEITPRSAKMNIQPRQA-----QAKKAQ-IFPERTLNP 1108
Qy 524 KGLPEFSQOEKSKAINQEGLE-----EDDNITERLSPENSPIOQENAGL 571
Db 1109 KKELIITNAKEALIAINQKGLDLHFIRDHFTTQSLQETTIKELLEOKLIYKDHONGGY 1168
Qy 572 GASPD-----KPYMKQVONQRYYLAKSOIQ--ELIKAKDYTKLAKLLSNRHTYNIS 621
Db 1169 ILANDYLSGNVVRKLEKVEKAINQVEGLEANVKDELILIPK----- 1211
Qy 622 LRLKEQLPDVN-PRIPSGRDIEKAFVLDKTEKKNQIYSSASPVFQNKWSLFGYRYL 680
Db 1212 LKATEIMANINSPIPT-----QYLEEFMLLSANHYEKQYDGKMTDYQ----- 1255
Qy 681 LGLDPKQTIHELVLGQKAGLQFEGYNLPSDFNLEDLKNIRITPLFSQDNFKLSLLD 740
Db 1256 ----LSNLKEDIKIEHLISG-----AYEVFVRNNELNELYGRHDKKPHSYKVPESLNNK 1306
Qy 741 FNNYDYGKIKAFGLPLFLPKELRRNSSGSGQNSNSPWEQBIISQFQDNLSNQ--D 798
Db 1307 VLNNKOLSVKQAQV-----DPNDPKKEIFITDEQSNLARQKAE 1345
Qy 799 QLAQ--FSTKIWEKIIGDNEFDQNNRLOYLKLLDQESWINKTRDNLVWYVLGDKLVKP 857
Db 1346 ELKEAFKDWIYK-----DYSRRTHLE-QIYNDTFNNVLKTYD-----GSQLELEG 1390
Qy 858 KN-----NLEAKFROISN-----LOELLTAFYTSALSNNWYVDSGAKSTIIFEEI 905
Db 1391 FNYHISLRPHQKNIAIFRTIQRAVCLDHQVAGKTLCAIASCMEQKRMGLVKNKTLI----- 1446
Qy 906 AELDPKVEKGVADYVQLKFHYAIGFDDNAGKFNQEVIRSSRTIYLTSGSKSLEADTI 965
Db 1447 -AVPHLTKQWGEFYKAYPNANVLVDSKOTTEKE-----RELLFNQIANNYDAVVI 1499
Qy 966 DQLOAVRNAPLQSLQSYLDTERFVGVOKLATSIAVQHKQKEKTLPKKLNDGYTLIHDK 1025
Db 1500 AHTHLELLSNPRGI-----IEELKEEELVNAEKNFERQELAYKN 1538
Qy 1026 LKKPVIPOISSSPEKDWEGKLNQSQNVNVSIFGSIIESPYSTFQEDADLDQDQG 1085
Db 1539 ----NPRETKKPNRPAKFKLDK-----IRAKY-----DAILEKQGS 1571
Qy 1086 --DDSRQGNISLDNOEAGLLK-----QKLAILLGNQ-----FI--QYVQNDK 1124
Db 1572 HIDSQMGIDNLIVDEAHLFKNLAFETSMEKIA-GLGNQOQSNRARDLFIKTRYLHONDK 1630
Qy 1125 EIEF----EIIINVEKVSLSFVPEFKLAKTLEDNKGKTRVLSDETMSLIVNTTIKTPEM 1180
Db 1631 KIMFLGTGTPIAN--SISEMYHLQRYLTPDVLKERGL----- 1664
Qy 1181 SAVPEVDFTKWBOY-----DPRTPLAKT-KFVLKFKDQIPVDGSGNISDKWLASIPLV 1234
Db 1665 ----EFPD--DMAKTYGEVNDPELDTSAQSYKVMNRFSKFSVQV-----LSTMYRAFADIV 1716
Qy 1235 IHQOQMLRLSP-----VVKTIRES-----LGLKTEQOQOQOQOQ--OQOQOQPKK 1274
Db 1717 SNDDILKGNPHFPVKYGVKDPINVVVVRSEEAQFIVALENGKYNESIIDRMQKCEGK 1776

Qy 1275 AVRKEBELESTYNPKDFENILNPLTKAHLRLTSL-NLVNNDPNYKIEDLKVKINEAGDHOLA 1333
Db 1777 KSQKQGD-----NILSCTTDDARKVALDYRLI--DPAKAVEK-----EFSKS 1815
Qy 1334 FSLRANNIKRLMNTPTIFADYNPFFYFYNEDWRSIDKYLANKGNVSSHQQAAGNQGSGL 1393
Db 1816 YAMAKNIYENYLETHATKGTQLGF1-----GLSTPKTHSQKVS----- 1853
Qy 1394 IQRLNKNIKEPTPTPALIAIKDRNN---TNLSNYSKIIIMIKPKYLVERSIGVPMSTGLD 1450
Db 1854 LEALDNAHETENKPL-----DKAQLLESLSYDEKGNLIAP----- 1891
Qy 1451 GYIGSQTKDGTSSSQKGFQDQDFIQAQLGKNTYHGKLGLSIRIFDPCNELAKIKDAS 1510
Db 1892 ----SKKELENELKEKAKSVNLDIEIAKGCSDVYSDVLRHLVQMGIPQNEIAFIHDA- 1946
Qy 1511 NKKEBEKLLKSYDLFNKYLNEYKK----SP-KIAKGWTNI-----HPDQ- 1550
Db 1947 --KTES---QKQDLFKK-LNRGGVRVLGSPAKMGVG-TNVQERLVAMHEDLDCPWRPDEL 1999
Qy 1551 ----KEYPNQKLPENYLNVL-----NOPW-----KVT 1576
Db 2000 LQMEGRGIROGNILHQNDPENFRMKIYRYATEKTYDSRMMQIIETKSKGIEQFRNAHKJG 2059
Qy 1577 LYNSSDFITNLVFEPEGSDRGSGTK-----LKQV-IQOVNNNYADWSAYLTFWY 1626
Db 2060 LNELEDF--NM-----GSSNASEKBAEATGNPLIIEVKLRARIKSEESKYKAFNKEHYF 2112
Qy 1627 DKNIITNOPVITANITADVFIDKVELED--NTKLIAPNITO-WMPNISGSKEKPY----- 1679
Db 2113 NEESLKN-----NASKLDYLBQELKOLETLQRSVITPTHTTEIKLYDLKNEESKDYELIKV 2167
Qy 1680 ---KPTVFFGNWENENSSMNSQAQPTWEEKIRE-----GFALQALKSSPDQKTRTF 1727
Db 2168 KEVEPL-----KENASMSSEEL---THKKLKSQNKQIAEQNKKEKLDIAIKQFASNLT 2217
Qy 1728 VLITNAPLPLMYGPGPGF-----QNGPNFKTQDMRLVFPND----- 1763
Db 2218 FVNEEDDYKLEY--KGFVNVNAKTKYQVFEFSLSPKDNFNIAVSPSNVYKNDTINMPSS 2275
Qy 1764 -----DNQIAALRVQEQDRPEKSSSE 1783
Db 2276 YNFCABIKFDGFLKRLDNATIKL-----PEKIKE 2304
RESULT 28
ID ADE56037 standard; protein; 3259 AA.
XX ADE56037;
AC ADE56037;
XX 29-JAN-2004 (first entry)
DT Human Protein CAA53052, SEQ ID NO 1876.
XX Human; pain; neuronal tissue; gene therapy;
DE spinal segmental nerve injury; chronic constriction injury; CCI;
KW spared nerve injury; SNI; Chung.
XX Homo sapiens.
XX WO2003016475-A2.
XX 27-FEB-2003.
XX 14-AUG-2002; 2002WO-US025765.
XX 14-AUG-2001; 2001US-0312147P.
PR 01-NOV-2001; 2001US-0346382P.
PR 26-NOV-2001; 2001US-0333347P.
XX (GHO) GEN HOSPITAL CORP.

Qy 1240 LRLSPVVTIRELGLKTEQQ-----QQQQQQQQQQP-----OKKAV 1276
Dy 3076 LR-----ENQHYGDLNHCVALEKQVQELQAGPLNDVAPGAPQKNGV 3120
Qy 1277 -RKBEELTYNPKDEFNINPLTKAHLTLNLVNDPNVYKIEDLVKIKNEAGDHQL--- 1332
Dy 3121 HRKSDPELRPEQSQFSEAQ-----QQLCNT--RQEVNELRKLEERDORVAE 3168
Qy 1333 -AFSLRANNIKRL-----MWTPI 1349
Dy 3169 NALSDVAERQIRRLHSEWDSRTPI 3193

RESULT 29
ID ADE56033
AD56033 standard; protein; 3259 AA.
AC ADE56033;
XX
XX
DT 29-JAN-2004 (first entry)
XX Human Protein CAA53052, SEQ ID NO 1872.
DE Human; pain; neuronal tissue; gene therapy;
KW spinal segmental nerve injury; chronic constriction injury; CCI;
KW spared nerve injury; SNI; Chung.
XX Homo sapiens.
XX WO2003016475-A2.
XX 27-FEB-2003.
XX 14-AUG-2002; 2002WO-US025765.
XX 14-AUG-2001; 2001US-0312147P.
PR 01-NOV-2001; 2001US-0346382P.
PR 26-NOV-2001; 2001US-0333347P.
XX (GEO) GEN HOSPITAL CORP.
PA (FARB) BAYER AG.
PA
PI Woolf C, D'urso D, Befort K, Costigan M;
XX WPI; 2003-268312/26.
DR GENBANK; CAA53052.
XX
PT New composition comprising two or more isolated polypeptides, useful for
PT preparing a medicament for treating pain in an animal.
XX
XX Claim 1; Page; 101pp; English.

CC The invention discloses a composition comprising two or more isolated rat
CC or human polynucleotides or a polynucleotide which represents a fragment,
CC derivative or allelic variation of the nucleic acid sequence. Also
CC claimed are a vector comprising the novel polynucleotide, a host cell
CC comprising the vector, a method for identifying a nucleotide sequence
CC which is differentially regulated in an animal subjected to pain and a
CC kit to perform the method, an array, a method for identifying an agent
CC that increases or decreases the expression of the polynucleotide sequence
CC that is differentially expressed in neuronal tissue of a first animal
CC subjected to pain, a method for identifying a compound which regulates
CC the expression of a polynucleotide sequence which is differentially
CC expressed in an animal subjected to pain, a method for identifying a
CC compound that regulates the activity of one or more of the
CC polynucleotides, a method for producing a pharmaceutical composition, a
CC method for identifying a compound or small molecule that regulates the
CC activity in an animal of one or more of the polypeptides given in the
CC specification, a method for identifying a compound useful in treating
CC pain and a pharmaceutical composition comprising the one or more
CC polypeptides or their antibodies. The polynucleotide or the compound that
CC modulates its activity is useful for preparing a medicament for treating
CC pain (e.g. spinal segmental nerve injury (Chung), chronic constriction

CC injury (CCI) and spared nerve injury (SNI)) in an animal (e.g. gene
CC therapy). The sequence presented is a human protein (shown in Table 2 of
CC the specification) which is differentially expressed during pain. Note:
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic form directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 3259 AA;
SQ
Query Match 2.8%; Score 276; DB 7; Length 3259;
Best Local Similarity 18.8%; Pred. No. 2e-06;
Matches 310; Conservative 240; Mismatches 505; Indels 590; Gaps 76;

Qy 106 IPKKYQVWISFSPEDDKERFLGFLHKEKLEQG-----NIAQSATKF 148
Dy 1738 LSKKFSQSLMEKSLSEVQ-----DLKHQIEDNVSKQANLEATEKHNDQNTVEEGTQS 1792
Qy 149 IYLLPLDMPKAALQVSYIVDKNPNLIHPLNFSQAQSIKPLALTRSSDFIAKLNPNN 208
Dy 1793 I-----PGETEEDSL-----SMSTRPTCSSEVPSAKSANPAVSKD-----FSS 1831
Qy 209 QDELWVLEKFFDLEALKANIRLOTADPFSEKGNLVDPFVYSFIRNPQNKQEWASDLNOD 268
Dy 1832 HDEINNYLQO---IDQLKERI---AGLEEEK-----QKNKFSQTLNE 1869
Qy 269 QKTVRLVLRTEFSPQAQKTIKDYKDETFELSSIDLKASNGTSL-----FANENDLKQOLD 324
Dy 1870 KNTLLSQISTK-DGELK-MIQSEVTKNLLNQIQEELSRVTKLKETAABEKDDLEERLM 1927
Qy 325 VDLDDVDFYFGQSETITS-----NSQVKPV-----PASERSLKDRVFKKDDQ 368
Dy 1928 NQLAELNGSIGNYCQDVTDAQIKNELLESEMKNLKCVCSELEEKQQLVKEKTVESIR 1987
Qy 369 KPRIEKF--SLYEVDALSFYSQLOELV-SKPNISKDLVN-----ATLARNLR----- 412
Dy 1988 KEYLEKIQGAQKEFGNKGSHAKELQELLKEKQEVKQLOKOCIRYQEKISALERTVKALEF 2047
Qy 413 -----FSLGKYNFLFDDLASHLDVYFLVSKAKIKOSS 445
Dy 2048 VQTESQKDLKITKENLAQAVEHRKKAQAEALASFVKLLDDTQSEA-ARVLADNLKXKELQ 2106
Qy 446 TKKLFIELPIK---ISLKSILGQEPNITKLPFEKVTFFKLDNFR---DVEIEKAFGLLYP 500
Dy 2107 SNKESVKSQMKQKDEDLERRLEQAEKHLKE--KKNMQEKLDA LRREKVLHEETIGTQV 2164
Qy 501 GVN---BELEQARK-----AQRASPEK-----EKSXKGLKES----- 530
Dy 2165 TLNKKDKVEQQLQENLDSVTVTQLAAFTKSMBSLQDRDRVIDEAKKWERKFSDAIQSKEE 2224
Qy 531 --QOKEENSKAINNQ-----BGLBEDDNITERLPENSPYQYQENAGLGA 573
Dy 2225 BIRKEDNCSVLKQDLQMSIHMEELKINIRLHEDHQIWEKQAQ-TEVQLQKVCYDITLQ 2283
Qy 574 SPDPKYMIKOVQORYYLAQSIQELIKAKDYTKLAKLLSNRHTYINISLRKEQLFDVNP 633
Dy 2284 GENKE-LLSQLEETRLHYHSSQ-----NELAKLESELKS-----LKDQLTDL-- 2324
Qy 634 RIPSRRDIEKAFVLDKTEKNKYQIYSSAPVQNKWSLFGYYRYLLGLDLPKQTIHEL 693
Dy 2325 ---SNSLEKCK-----BQGNLEGITIRQ----- 2344
Qy 694 KLGQKAGLOFE--GVENLPSPFNLE-----DLKNIRIKTLPFSQKDNFKSLD 740
Dy 2345 ---QEADIQNSKFSYEQLEDTQASRELTSRLHEINKEQKIISLLSGKEAIOVAE 2401
Qy 741 FNNYDGIKAIKAFEGFLPLFKELRRN-----SSNSGGSQNSNPWEQIISQFKDQNL 796
Dy 2402 LRQHQDKIKE-----LENLSQEEENIVLEENKKAVDKTNQ--LMTLTKIKKENIQ 2455
Qy 797 QQQLAQFSTKI-----WEKIIIGDNEFD-----QNNRL----- 824
Dy 2456 KAQLDSFVKSMSSLLQDRDRIVGDYQQLERHLSIILEKQDIQBAEAENKLEIRGL 2515

Qy	825	-----QYK-----	-----LLKQLQBSWI-----	NKTRDNLVWTVLGD	851	FT	/note= "one of 21 complete segments of homology of 22
Dd	2516	RSMDLLNSENAKLDAELIQVREDLNQVITIKDSQKQLLEVQLQNKLENY-AKLBE	2574	873	FT	amino acid length"	
Qy	852	KLKVKPKNN--LEAKF-----	-----RQISNLOE-----	873	FT	/label= tandem repeat	
Dd	2575	KLKSEBANDLRRSFNALQBEKQDLSKBIESLKVSIQSLTRQVTAQBEGTGLYHAQL	2634	873	FT	/note= "one of 21 complete segments of homology of 22	
Qy	874	-----LLTAFYTS-----	-----AALSNNWYQDSGAKSTIIFEBIAELDPKVKKEVGADV	920	FT	amino acid length"	
Dd	2635	KVKEEVHRLSALFSSSQKRIAELEELVCQKEAAK-----	-----KVGEIEDKLK-----	2682	FT	/label= tandem repeat	
Qy	921	YOLK-FHYAIGFDNAGKFNQEVIRSSRTIYKTSKSKLEADITDQLNOAVNNAPLGL	979	FT	amino acid length"		
Dd	2683	-ELKHLHDAGIMRNETETAEERVAELARDIV-----	-----EMEQLKL-MVTKENKGLTAQI	2733	FT	/label= tandem repeat	
Qy	980	QSF-----	-----YLDTERFGVFOKLATSLAVQHKQEKTKPKKLNNDGYTLIHDKLK	1027	FT	/note= "one of 21 complete segments of homology of 22	
Dd	2734	QSGRSMSSLQNRDHANEELDELKRYDASLKELAQKEQGL-----	-----LNRE-----RDALL	2785	FT	amino acid length"	
Qy	1028	KPVIQISSSPKDWFE-GKLNQNGSQNVNVSFTGSIIESPY-----	-----FS-----TNFOEDA	1078	FT	/label= tandem repeat	
Dd	2786	SETAFSMNSTEENSLHLEKLNQQLSKDQELHLSQLEDSYNQVSFKAWASLQNER	2845	FT	amino acid length"		
Qy	1079	D-----	-----LDQDQDGS-----	1114	FT	/label= tandem repeat	
Dd	2846	DHLWNELEKPKRSEBQKRSAAQFSTSPAERVSQSLKAMSSQLNDRDRLLKE-----	-----LKNL	2900	FT	/note= "one of 21 complete segments of homology of 22	
Qy	1115	FIQYQON-----	-----DKEIFEIINVE-----	1135	FT	amino acid length"	
Dd	2901	QQOYLQINQIEITELHPLKAQOEYQDKTKAFQIMQBELROENLSWQHLEHQLRMEKSSWE	2960	FT	/label= tandem repeat		
Qy	1136	VSELSFRVEFKAKTLEDNG-----	-----KTIRVLSDETMSLIYNTTIEKTPEMSAYPE--	1185	FT	/note= "one of 21 complete segments of homology of 22	
Dd	2961	IHERMKQYLMASDDKQQLSHLQNLIRELRSSSQTPQLKVQYQRAQSPETSASPDGS	3020	FT	amino acid length"		
Qy	1186	---VFDTKWVQYDPTPLAAKTKFVLKFK---	---DQIPVDGSGNISDKWLASIPLVTHQOM	1239	FT	/label= tandem repeat	
Dd	3021	QNLVYETELL-----	---RTQLNDSLSKEIHQKELRIQQLNSNFSQLLEEKNTLSIQLCDTSQS	3075	FT	/note= "one of 21 complete segments of homology of 22	
Qy	1240	LRLSPVVKTIRELGLKTEQQQ-----	-----QQQQQQQQQQP-----	1276	FT	amino acid length"	
Dd	3076	LR-----	-----ENQOHGYDLLNHCAVLEKQVQELQAGPLNIDVAPGAPQKNGV	3120	FT	/label= tandem repeat	
Qy	1277	-RKBEELTYNPKDEFILNPLTKAHLRLTSLNVLNNDPNYKIBDLKVIKNEAGDHQ---	1332	FT	amino acid length"		
Dd	3121	HRKSDPEELRPPQSFSEAQ-----	-----QQLCNT--RQEVNELRKLLEERDQRVAE	3168	FT	/label= tandem repeat	
Qy	1333	-AFSLRANNIKRL-----	-----MNTPI	1349	FT	/note= "one of 21 complete segments of homology of 22	
Dd	3169	NALSVABEQIRRLSEHSEWSSRTPI	3193	FT	amino acid length"		
AC	AAR46608;						
XX	RESULT 30						
ID	AAR46608						
XX	AAR46608 standard; protein; 1663 AA.						
XX	AAR46608;						
DT	16-OCT-2003 (revised)						
DT	25-MAR-2003 (revised)						
DT	22-SEP-1994 (first entry)						
XX	Plasmodium falciparum erythrocyte membrane protein PfEMP3.						
DE	Plasmodium falciparum erythrocyte membrane protein; PfEMP3; malaria;						
KW	antigen; epitope; vaccine; anti-idiotypic antibody.						
XX	Plasmodium falciparum; (Malayan Camp strain).						
OS	Key						
FH	Location/Qualifiers						
FT	472..493						
FT	Region						
FT	/label= tandem repeat						

XX	15-SEP-2003; 2003WO-US028789.	QY	407	LARNLFSLGKYNFLPDDLASHLDYFLVSVKAKIKOSSITKXLFIELPKISIKSSILGD	466
XX	13-SEP-2002; 2002US-0410303P.	DB	3472	LANNPDVNAIRTN-----KVETEQINNEL-----TQAKQGLTVD	3505
XX	(TEXA) UNIV TEXAS A & M SYSTEM.	QY	467	QEP--NIKTLFEK-----EVTFKLONFRDVEIEKAFGLLYPGVNEBLEAR-K	511
PA	(INH1-) INHIBITEX INC.	DB	3506	KOPLINAKTALQOOSLDNOPSTTGMTTEATIQNYAKRQKAEQ-----VIQNAKIIENAOPS	3561
XX	(UABR-) UAB RES FOUND.	QY	512	AGRASPEKSKKGLKEFSQ-----OKEENSALNNQGLEEDDNIATERLPENSPI--	562
PI	Hook M, Xu Y, Sillanpaa JV, Sthanam N, Ponnuraj K, Patti JM;	DB	3562	VQOVSEKSKVEQALSELNNAKSALRADKQELQAYNQ---LIQPTDLNNKFPASITAYN	3618
XX	WFI; 2004-315684/29.	QY	563	-YQOENAGLGAS-----PDKPYMIKDVQORYYLAQSOIQELIKADYTKLAKLUS	613
XX	Identifying LPXTG-containing cell wall-anchored surface proteins from	DB	3619	QRYQQPSNELNKTNTDRILKEQNPVSADVNN-----ALNKVREV--OQKLNARALLQ	3671
PT	Gram positive bacteria, for treating infection caused by the bacteria,	QY	614	NRHTYNISLRKLEOLFVNPRIPSS-----RDIKAKFVLDKTEKN	654
PT	comprises searching sequence information database for the sequence having	DB	3672	NKEDNSALVRAKEQLQQAQVDPFSTEGMTQQTQKDDYNSKQAAQOQISKAQVINDGAT	3731
XX	LPXTG-motif.	QY	655	KYWOIYSSASPVFQNKWSLFGYRYLLGLDPKQTIHELVLGQKAGLQEGYENLPSPDN	714
PS	Claim 16; SEQ ID NO 23; 96pp; English.	DB	3732	T--QQISNAKTUVERALEALNNAKTGLRADKEELQAYNQLTQ---NIDTSGKTASIR	3785
CC	This invention relates to a novel method of identifying LPXTG-containing	QY	715	LEDLNIRIKTLPFSQKONFKLSLLDFNNYD-----GEIKA--PEFGLPLPLPKELRR	766
CC	cell wall-anchored surface proteins from Gram positive bacteria that bind	DB	3786	KYNEAKSRIQTQIDSAKNEANSILTNDNPQVSOVTAALNKIKAVQPELDKAILMKNKEN	3845
CC	to an extracellular matrix molecule which comprises searching a database	QY	767	NSNSGGSQ-----NSNSPWE---QEIISQDKQNSLQDOLAQFSTKIWEKII--GDE	815
CC	of sequence information for a putative protein sequence having the LPXTG-	DB	3846	NNALVQAKQOQLOQVNEVDPTQGMTTDTANNYKSKKREADEIQKA-----QQIINNGDA	3900
CC	motif in its C-terminal region and analysing the sequence for the	QY	816	NE---FDQNRRLQYKLLKDLQOSWINKTRDNLTYLWYLGDKLVKPKP-----NNLEAKFRQ	867
CC	presence of one or more immunoglobulin (Ig)-like fold regions. The	DB	3901	TEQQTINETNRVNQAI-----NALNKAQNDL---RADKSOLEWAYNQLIQNVDTNGKK	3950
CC	invention may be useful for the production of compounds with an	QY	868	ISNLQELTAFYTSAAALSNNMYQDSGAKSTIIFEEIAELDPKVKKEKVA-----DYOQL	923
CC	antibacterial activity or for production of a vaccine. In addition the	DB	3951	PASIQOYQAA---RQAIETQYN-----NAKSE-AHQILENSNPSVNEVAQALQKVEAVQL	4001
CC	disclosed sequences may be useful for gene therapy. The antibody is	QY	924	KHYAIGFDDNAGKNQEVIRSSRTIYLTSGSKLEADTTIDQLN-----QAVK	973
CC	useful for treating or preventing an infection of Gram-positive bacteria	DB	4002	KYNDAIHILQN---KENNSALVTAKNLOQOSVNDQPLTTGWTQDSINNYEAKNEAQAIR	4059
CC	in a human or animal patient. The method and the proteins are useful in	QY	974	NAPGLQ-----SFYLDTERFGVFOKLATSLAVOHKQKEKTLPKLKNNDGYTLIHDKLKK	1028
CC	generating antibodies for treating and preventing the spread of	DB	4060	NAEAVINNGDATAKQISDEKSKVEQALA-----HLNDAKQQLTADTTTEL	4103
CC	infections of Gram positive bacteria, for interfering with, or inhibiting	QY	1029	PVPIQISSPEKDWPEGKUNQNGQSONV---NVSTFGSIIESPYFSTNFQEDADLDQDQG	1085
CC	binding interactions by Gram positive bacteria, for monitoring the level	DB	4104	QTAVQ-----QLNRRGDTNNKPRSRINAYNKAIOQL-----4134	
CC	of gram positive bacterial antigens, or antibodies recognising the	QY	1086	DSROGNSLNOEAGLLKQKLAILLGNQFIOVYQONDKEIEFEIINVEKVS-----L	1139
CC	antigens in a human or animal patients suspected of containing the	DB	4135	ETQITSANANAVIQPIRTVOEVNNAQOVNQLNQLTEALNQLFSLNNDALKAA	4192
CC	antibodies, in preventing or reducing infection of medical	QY	1140	SFRVFPEKLAKTLEDNGKTIIRVLSDETMSLIVNTTIKTPMSAVPEVFTKWEQYDPR	4199
CC	devices and prostheses caused by such organisms, and in treating or	DB	4193	RLNLENKINQVTDGWT-----QOSIEAYQAKRAQVNESNT	4230
CC	preventing infections in highly susceptible groups such as premature	QY	1200	PLAAKTKFVLKPKDQIPVDGSGNISDKWLASIPLVHQMLRLSPVVK---TRELGLKTE	1257
CC	newborns, AIDS and debilitated cancer patients, and bone marrow	DB	4231	ALA-----LINNGDADEQQITTETDRVNOQTTLNLTQAINGLTVNKEPLET-	4275
CC	transplantation. The present sequence is that of a surface anchored LPXTG	QY	1258	QOQQOQQOQQOQQOQKKAUKRKEELETYNPKDEFNLNPLTKAHLRLLSLNVLNNDNYKI	1317
CC	protein identified using the method of the invention.	DB	4276	AKTALQNNIDQVPSTGDM--TQOSVANYNQKLQ-----IAKNEINTINNVLANNP	4323
XX	Sequence 10203 AA;	QY	1318	EDLKVIK-NEAGDHQLAFSL-RANN-----IKRLMNTPI---TFAD---YNPFF	1358
XX	Query Match 2.8%; Score 275.5; DB 8; Length 10203;	DB	4324	DVNAIKTKNAEAEIRISNDLTQAKNNLQVDTQPLEKIKRQLQDEIDQGTNTDGMTQDSVD	4382
XX	Best Local Similarity 18.5%; Pred No 1.2e-05;				
XX	Matches 375; Conservative 319; Mismatches 758; Indels 573; Gaps 93;				
QY	74 LNGBT-----FDPKSEFTDFVSKFDFLTNNGRTVLEIPKQYQVVISFSPEDDKERPR	127			
DB	3189 LMGWTQESINNYTKRREAQNIASSADTIINNGDASIE-----QITEN-----	3231			
QY	128 LGFHLKEKLEDGNIAGSATFYVLLPLDMPKALGQYSYVDKN-----FNNLIH	178			
DB	3232 -----KIRVEATNALNEARQHLTADTTLSLKTVEKLSRRGDTNNKPSVSAYN--TIH	3285			
QY	179 PLNFSQAQSTKPIALTRSSDFIAKQFNQDQELWVLEKFFDLKALKANIRLOTADSF	238			
DB	3286 SLOSETQT--ENRANTIINPIRSVEVNALH-----EVNQLNQRUT-----	3327			
QY	239 EKNLVDPFVYSFIRPNQKQWASDL-NODQKTVRLYLRTFSPQAKTILKDYKYKDET	297			
DB	3328 DTINLLQP-----LANKESLKEARNRLESKINETVQTDGWTQS-----VENYK	3371			
QY	298 FLSISDLKASNGTSL-----PANENDLKDQ-----LDVLLDVSYFGQSTIYNSQ	346			
DB	3372 ---QAKIKAQNESSIAQTLINNGDASDQEVSTIEKLNQKLSLTN-----SINHLTVNKE	3424			
QY	347 VKPVPASERSLKDRVKFKDQKQPRTEKFSLEYDALSFSYQQLVELVSPNSIKDLWNAT	406			
DB	3425 --PLETAKNQLQANI-----DQKPSD--GMTQOSVQSVYERKLOEAKDKINSI-----NNV	3471			

QY 1359 YNEDWRS-----IDKYLNNKGNVSHQOQAGGQSGGLIQ----- 1395
D 4383 NYNDSLSAAIEKGKVNKLRNPTVEQVESVANAQQ-----VIOQLNARTSLVDPKTQL 4439
QY 1396 -----RLNKNIKPPTFPALTALKDRNNTLSNYSKIIIMIKPKYLVERSIGVPMWSTGLD 1450
D 4440 QEAKRLENSINQDTTDCM-----TQDSLNNYNDK--LAKARONLEKI-----SKVLG 4486
QY 1451 GYIGSEQTKGTSSSSQKGFDDPIQALGLK-----NTEYH-----GK 1489
D 4487 GQPTVAEIRONTDEANAHK-----QALDTARSQLTNRBPYINHNESHNLNAQKDN 4539
QY 1490 LGLSIRIPDPNGLAKIK-----DASNNKGKELL-----KSYD- 1523
D 4540 FKAQVNSAPNHTLETIKNKADTLNQSMTALSESIADYENQKQENYLDASNNKRODYDN 4599
QY 1524 ---LFFKNYLN-----YEKSPKIAKGWTNIHPDQKEYPNPKLPENYLNILV-- 1568
D 4600 AVNAKGLNLTQPTMSADVIDQKAEDVKTKTALDGNQRLVAKQOAL--NHLNTLND 4657
QY 1569 LNQPWKVTLNYSDFITNLFVEPEGSDRGSGTKLVQIKQVNNNYAD--WGSAYLTFWY 1626
D 4658 LNDARQRTLTDTINHSNPNINSVQAKKANTVNTAMTQTKOTIANYDDELHDGNYINADK 4717
QY 1627 DK-----NIITNPQVITANTADVFIDKVELEDNTKLIAPNITQWPNISGSKEFKYP 1681
D 4718 DKDAYNNAVNNAKQELINQSDANQAQLDPAINKVYTORVNTTKNDLNGNDKLABAKRDAN 4777
QY 1682 TVFFG-----NWNEN-----SSMNSOAOQPTWEKIREGPAQALKSSFD-----Q 1722
D 4778 TTIDGLTYLNEAQRKAKENGVKASTKNTITSQLODYNLQ--NIAMQALRNSVNDVNVK 4835
QY 1723 KTRTPVLTNAPLPLWKYP-----LGFQNGP-----NFKTDWRLVFNQDNDQIAA 1769
D 4836 ANSNYNEDNGPKAYNAVQVTHAQTFLINAQSNPEMSRDVWVQKTOQAVNTAHQLHGQOKL 4895
QY 1770 LRVQOQ-----DREKSESDKQKWKIKFKVVIPEEMFNSGNIR 1808
D 4896 EQAQQSANTTEIGNLPLNTQKAKEK-----ELVNSKQTR 4930
RESULT 33
ADSO4803
ID ADSO4803 standard; protein; 10203 AA.
AC ADSO4803;
XX
XX
DT 04-NOV-2004 (first entry)
XX Staphylococcus epidermis polypeptide seqid 4098.
DE antibacterial; vaccine; antisense therapy; Staphylococcus epidermidis;
KW recombinant expression vector; infection; computer readable medium;
KW computer based system.
XX Staphylococcus epidermidis.
XX US2004147734-A1.
XX
XX 29-JUL-2004.
XX
XX 01-DEC-2003; 2003US-00724972.
XX
XX 08-NOV-1997; 97US-0064964P.
PR 13-AUG-1998; 98US-00134001.
PR 29-NOV-1999; 99US-00450969.
XX
XX (DOUC/) DOUCETTE-STAMM L.
PA (BUSH/) BUSH D.
XX Doucette-Stamm L, Bush D;
XX
XX WPI; 2004-580138/56.
DR

DR N-PSDB; ADS01031.
XX New isolated polypeptide and encoding nucleic acid derived from
PT Staphylococcus epidermidis, useful for diagnosing, preventing and/or
PT treating an S. epidermidis bacterial infection.
PS Claim 17; SEQ ID NO 4098; 741pp; English.
XX
CC The invention describes an isolated nucleic acid comprising a nucleotide
CC sequence with any of 3772 fully defined nucleotide sequences (SEQ ID NO:
CC 1-3772) and encoding an Staphylococcus epidermidis polypeptide with any
CC of 3772 fully defined amino acid sequences (SEQ ID NO: 3772-7544) as
CC given in the specification. Also described are: a recombinant expression
CC vector; a cell comprising a recombinant expression vector of (1);
CC producing an S. epidermidis polypeptide; an isolated nucleic acid
CC comprising a nucleotide sequence of at least 8 nucleotides in length; a
CC vaccine composition for prevention or treatment of an S. epidermidis
CC infection, comprising a nucleic acid cited above and a carrier; treating
CC a subject for S. epidermidis infection; a recombinant or substantially
CC pure preparation of an S. epidermidis polypeptide or its fragment; a
CC vaccine composition for prevention or treatment of an S. epidermidis
CC infection; detecting the presence of a Staphylococcus nucleic acid in a
CC sample; a computer readable medium having recorded in it the nucleotide
CC sequences with SEQ ID NO: 1-3772 or its fragments; a computer based
CC system for identifying fragments of the Staphylococcus genome of
CC commercial importance; a computer based system for identifying fragments
CC of the Staphylococcus plasmids of commercial importance; identifying
CC commercially important nucleic acid fragments of the Staphylococcus
CC genome and/or plasmids; and identifying an expression modulating fragment
CC of the Staphylococcus genome and/or plasmids. The methods and
CC compositions of the present invention are useful for the diagnosis,
CC prevention and/or treatment of an Staphylococcus epidermidis bacterial
CC infection. This is the amino acid sequence of a S. epidermis protein of
CC the invention.
XX
SQ Sequence 10203 AA;
Query Match 2.8%; Score 275.5; DB 8; Length 10203;
Best Local Similarity 18.5%; Pred. No. 1.2e-05;
Matches 375; Conservative 319; Mismatches 758; Indels 573; Gaps 93;
QY 74 LANGKT-----FDPKSEFTDFVSKFDFLTNGRTVLEIPKKYQVWISFSPEDDKERFR 127
D 3189 LANGTQESINNYTTKREAQNIASSADTIINGDASIE-----QITEN----- 3231
QY 128 LGPHLKEKLEDGNTAQSAATKFIYLLPLDMPKAAAGQYSYIVDKN-----FNLLIH 178
D 3232 -----KIRVEATNALNEAKQHLTADTTSLKTEVRKLSRRGDTNNKPPSSVSAYNN-TIH 3285
QY 179 PLSNFSAQSIKPLALTRSSDFIAKLNQFNQDELWVYLEKFFDLEALKANIRLQADPSF 238
D 3286 SLQSEITQT-ENRANTIINKFIRSVENVNNAH-----EVNQLNQRLT----- 3327
QY 239 EKGMLVPFPVYSFIRNPQNKQEWASDL-NQDKTVRLYLRTEFSPQAKTILKDYKYKDET 297
D 3328 DTINLLQP-----LANKESLKEARNRLESKINETVQTDGMTQQS-----VENYK----- 3371
QY 298 FLSSIDLKASNGTSL-----FANENDLKQ-----LDVLLDVSDFGGQSEITNSQ 346
D 3372 ---QAKIKAQNESSIAQTLLNNGDASQDEVETEIKLNQKLSLNTN-----SINHLTVNKE 3424
QY 347 VKVPASERSIKDRVKPKDQKPRIKFSBYEYDALSFSYQLOELVSKPNSIKDLVNAT 406
D 3425 ---PLETAKNQLQANI-----DQKPSD--GMTQSQVSQYERKLQAKDKINSI-----NNV 3471
QY 407 LARNLRFSLGKYNFLFDDLASHLDYFVLVSKAKIKQSSITKKLFIEPIKLSKSIIGD 466
D 3472 LANNPDVNAIRTN-----KVETEQINNEL-----TQAKQGLTVD 3505
QY 467 QEP--NIKTLPFK-----EVTFKLDNFRDVEIEKAFGLLYPGVNEELQAR-K 511
D 3506 KOPLINAKTALQOQLNDNQFSTTGMTTEATINYNAKRQKAEQ-----VIONANKIIEAOPS 3561

QY 1205 TKFVLKFDQIPVDGSGNISDKWLASIPVIHQMLRLSPVKTIRELGKTKTQOQOQOQ 1264
 Db 1272 TSC-----LKETQDQLLES 1285
 QY 1265 QOQOQOQKKAARKEELETYNPKDEFNINPLTKAHLRLTSLNVDNPNYKIEDL-KVI 1323
 Db 1286 QKEKQLOEEAALSGELQ-----QVQENGDIKDSLKVKEELVKVL 1327
 QY 1324 --KNEAGDHQAFSLRANNIKRLMNTPIITFADYNPPFYNNEDWRSIDKYLNNKGNVSS-- 1379
 Db 1328 EEKLOAATSQIDAQATN--KELQELLV-----KSOENEGNLQGES 1366
 QY 1380 -----HQQAAGNQGSLIQRLN--KNIKPFTFPALIALKORNTNL-----SN 1423
 Db 1367 LAVTEKLQLEQANGELKEALCKENGKELQGLK-----DESNIVLESOKSHNE 1417
 QY 1424 YSDKIIMKPKYLVERSIGVPWSGLDGYIGSEQTKDGTSSSQKGFQDFQFQALGLKN 1483
 Db 1418 IQDKLEQAQK--ERTLOEETSKLAELQSLQKQANEELQKSLQK-----QLLLEKG 1467
 QY 1484 TEYHGKGLSIRIFDPGNEIAKIDA 1509
 Db 1468 NEPTQLAEYQKVIDEMDAAASVKA 1493

RESULT 36
 ID ABB61173
 XX ABB61173 standard; protein; 1690 AA.

AC ABB61173;

DT 26-MAR-2002 (first entry)

DE Drosophila melanogaster polypeptide SEQ ID NO 10311.

XX Drosophila; developmental biology; cell signalling; insecticide;
 KW pharmaceutical.

XX Drosophila melanogaster.

XX WO200171042-A2.

XX 27-SEP-2001.

XX 23-MAR-2001; 2001WO-US009231.

XX 23-MAR-2000; 2000US-0191637P.

XX 11-JUL-2000; 2000US-00614150.

XX (PEKE) PE CORP NY.

XX Venter JC, Adams M, Li PWD, Myers EW;

XX WPI; 2001-656860/75.

XX N-PSDB; ABL05276.

XX New isolated nucleic acid detection reagent for detecting 1000 or more
 PT genes from Drosophila and for elucidating cell signalling and cell-cell
 PT interactions.

XX Disclosure; SEQ ID NO 10311; 21pp + Sequence Listing; English.

XX The invention relates to an isolated nucleic acid detection reagent
 CC capable of detecting 1000 or more genes from Drosophila. The invention is
 CC useful in developmental biology and in elucidating cell signalling and
 CC cell-cell interactions in higher eukaryotes for the development of
 CC insecticides, therapeutics and pharmaceutical drugs. The invention
 CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
 CC sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-
 CC ABB72072). The sequence data for this patent did not form part of the
 CC printed specification, but was obtained in electronic format directly
 CC from WIPO at ftp.wipo.int/pub/published_pct_sequences

XX

SQ Sequence 1690 AA;
 Query Match 2.8%; Score 273.5; DB 4; Length 1690;
 Best Local Similarity 18.8%; Pred. No. 1e-06;
 Matches 321; Conservative 263; Mismatches 559; Indels 563; Gaps 76;
 QY 15 AITGSTVF-----GTVV-----GLASKVKYRGVNPQTQGVISQGLIDSVAF----- 55
 Db 140 AFIGDTHFAAGWAGVLDPEPKNKGCVSGKYFQCEPKRGIFSRLLTTLTYPLAGAQT 199
 QY 56 -----KPSIANFTSDYQSVKALL-----NGKTFDPKSSSEFTDFVSKFDELFTNN 99
 Db 200 PTSPLAKSPDRSRTVPTASIRSSMLRSPGIGKNGMAVGDRVIVSSGFGSPGLIRYL 259
 QY 100 GRT-----VLEIPKYOVIS-----EFSPEDDKERF-R 127
 Db 260 GETQFAPGNWCGVELDEPBGKNGDTVDIRYFECKPKYGVFVPIAKVUSLSPSSKKTLSR 319
 QY 128 LGPHLKEKLED-----GNIAQSATKFIYL-----LP-----LDMPKAALQGVSYIVDKN 171
 Db 320 TG--SRSLTSIGTWNISIATTATSRMRNABQKRSSTPVKPILATPKSQFSMDLLREKQ 377
 QY 172 FNNLIHPLSNFSAQSIKPIALTRSSDFTAKLNQF-----NNQDELWVYLEKFPDLEAKA 227
 Db 378 -----QHVEKLMVERDLDRDAQNQALQKQINEL-----KARIVELESALD 420
 QY 228 NIRLOTADF--SPEKGNLVDPFVYFIRNPQNO--KEWASDLNQDKTVRLYLRTFSPQ 283
 Db 421 NERKTEELQCSIDEAQFCGDEL-----NAQSVYKEKIHDL--ESKITKLVSAT---PS 470
 QY 284 AKTILKDYKYKDETFLSSIDLKASNGTSLPANENDLKQDLDVLDLVDVSYFGGQSEYITS 343
 Db 471 LQSILPP-----DLPSDDG-ALQEETAKLQEKWTIQQKEVESIAEQLE--- 513
 QY 344 NSQVQVPASERSLKORVKFKDQ-----QKPRIEKSLEYDALSFSYQLOELV 393
 Db 514 -----BEORLRENVKYLNEQIATLQSELVSQDALEKFLSECGIENLREL-ELL 563
 QY 394 SKNSIKDLVNATLARNLRFSLGKYNFLPDBLASHLDYVFLVSKAKIKQSISITKLFIEL 453
 Db 564 KEENE-----KQAEQAQAEFTKRL-AEK 585
 QY 454 PIKISLKSSILGDOEPNITLFEKVFVKLB-----NFRDVEIEKAFGLLYPGVNE 504
 Db 586 SVEVLRSLSELQNLKATSDSL-ESERVNKTDECEILLQTEVVRMRDEQIRE-----LNQ 636
 QY 505 ELEQ-----ARKAQRASFE-----KEKS-----KGLKEFSQOKEENKAINQEG 545
 Db 637 QLOEVTTQLNVQKADSSALDMLRLQKEGTEEKSTLLEKTEKELVQSQEAQAKTINDKEQ 696
 QY 546 LEEDDNITERLPENSP-IQYQOENAGLQASDPKPYMIKDVQNRVYLAISOIELIRAKD 604
 Db 697 LEKQISDLKQLAEQEKLVRENTENA-----INQIQ-----LEKESIEQ----- 734
 QY 605 YTKLKLKLSNRHTYNIISRLKEQLFDVNPRIPSSRDIEKAKFVLDKTEKNKYWOIYSAS 664
 Db 735 -----QALKQNELEDFQKQSESEVHLQEIKAQNTQKD--PELVESG- 775
 QY 665 PVFQNKWSLFGYRYRILGLDPKQTIHELKVLQVAGLQFEGYENLPDFNLEDKNTRIK 724
 Db 776 -----ESLKKLQOOLEQKTLGHEKLAQ--ALEELK--KEK 806
 QY 725 TPLFSQDNFKLSILLDFNNYVDGIEKAPFGLPLFLPKELRRNSSGGSQNSNSPWEQE 784
 Db 807 ETIIEKEQEELQQLQSKSASESALKVVQVQL-----EQLOQQAASGEESKTVAKLHD 861
 QY 785 IISQFKQNLNSQDL-----AQFSTKIWEKIIGD-ENEPDQNNRLOYLKLLKDLQESWI 837
 Db 862 EISQKQAEQTSQELKSTQSNLEAKSKQLEAANGSLEEEAKKSGHLLQITK--LKSEV 919
 QY 838 NKTRDNLVWTVYLGDKLVKPKKNLEAKPRQISNLQELLTAFYTTAAALSNWNNYQDSGAK 897
 Db 920 GETQAALSSCH-----TDVESKTKQLE-----AANAALKEKVNKVAESRAE 960

Db	392	DISNY-----NTEMLEI---SETLKVHKNFK-----	415
Qy	389	LQEL-VSKPNSIKOLVNATLARNRFSGLGKYNFLDLDLASHLDYYPVLSKAKIKQSSITK	447
Db	416	-OHLAVLDENNSTFVNNVLLKNINY--KKN---DOLIEGGE-----K	452
Qy	448	KLFLFELPKISLKSILGDQEPNITKTLFEKVFPLKDNFRDVELEKAFGLLYPCVNEELS	507
Db	453	KSFINL---INVD-----CYSSNSRLE	473
Qy	508	QARQAORASPE---KEKSKGLKEFQOKEENSAINNOEGLBEDDNIETLSPENPIQ	563
Db	474	NDENIERGKINMFTNDEKSNININNNNNNNNN--NNDNNMDNDVIIHKNKNNNIY	532
Qy	564	YQENAGLGASDPKPYMIKOVNORYYLAQSIOELIKAKDYTKLAKLLSNRHTYINSLR	623
Db	533	-----DNKYNVE-----CSSEKINDNGISNKNINILEPLNNLDTSNIFLE	572
Qy	624	LKEQ--LFDVNP---RIPSSRDIEKAKFVLDKTEKNYKWOIYSSASPVFQNKWSLFGYYR	678
Db	573	GKDEYKYVYVKNKEIRIPLFKEINK--IFEKPLKLYQLQD-----IKEEWTDNRIK	626
Qy	679	YLLGLDPQTHLVLKQKAGLQFEGY-ENLPSPDFNLEDLK-----NIR-IKTPLFQSK	731
Db	627	AIKSKDDMVFSQ-----VQLETVMRIKTDFFEKLKIKMAENIQSVGELLINK	677
Qy	732	DNFK-----LSLLDFN-----NYDGEIKAPFGLPL-758	
Db	678	DLSKNTDINIKDYNVLQKKSKKKKFLNDILNTVNTFTTESKYQDLYVKGESKEDIK	737
Qy	759	-----FLPKELRNS-----SNSGGSQNSPWEQELISQFDQNLSN	796
Db	738	NQIDFVTQECYRNNDIIRDTHKSDIFKNIKIDNNKYEIYNLEQEEINEKKNYKNN	797
Qy	797	QDLQAQSTKWEKIIGDENEF-----DONNRLOYKLLDQESWINKTRDNL-----	844
Db	798	NDSNKTFFLKI-----ENEFKDLLDSDQIFGDSLLADIKE---YNYTADNLDNNEN	848
Qy	845	-----YWT-----YLG-----DKLKVKPKNNLEAK--864	
Db	849	KSLYEDGENFITRNEPITNEYEKNIIYISDEQKYNEEDIIFKDKIKEKEKNNDTSD	908
Qy	865	FRQISNQLLTAFTYLSAALSNNWYQD-SGAKSTIIFEEIAELDPKVKGVADYVQL	923
Db	909	FENCIS-VQEKI---YVNEKIEYNNKNDKSSSSIIILEIKYKKEKDELSVFNLCVL	964
Qy	924	--KFHYAIGFDNAGKFNQEVIRSSSRITILKTSKSKLEADTIDQLNQAVKNAPLGLQS	981
Db	965	LDEFHSNDLENN-----YISVSSD-----DMKTNVSKNNITGVKE	1000
Qy	982	FYLDTERFQVFOK-----LATSLAVOHKQKKTLPKKNL---NDGYTLIHKLKPVP	1033
Db	1001	NKVDTNVEYDKGDDGVIEISFDSHKLKESKFDNNNIYDND-----DELEKNLSK	1054
Qy	1034	ISSSEKDWFEKLN-QNGQSQNVNVSFTGSIIESPYSTNPFQEDADLDQDQDSDRQGN	1092
Db	1055	YISDVKNNHNNIYNIEREDEREN-----EFVENKIQST-----ESHKSN	1095
Qy	1093	NSLDMQEAELLQKLA-----ILGNQFOIYQONDKEIETEPINNVKVSLSFRVE	1144
Db	1096	EFICTENKSLRQOYMSKEDISNRILKSDDDINLSKQN---YFELLDDKKQVMDNFQNN	1151
Qy	1145	F-----KLAKTLEDNGKTIIRVLSDEMTSLIVNTTIKTEPMSAVPEVFTKWEQYDERT	1199
Db	1152	IEQNNDKLAKEDKLDEGAYFEYLED--NKIIDSYIKETNKE-----NEELIKEY---	1197
Qy	1200	PLAAKTKFVLKFKQDQIPVDSGNSIDSKWLASI-----PLVHQMLRLS-----PV	1245
Db	1198	-----KKLK-----KNNTEINDEMNDIDKLLNFFGIPYIQSPCEAAQCSYLNKNKYCAI	1249
Qy	1246	VKTIRELGL---KTEQQQQQQQQQQQPKKAVRKEELETYNPKDEFNI-----LNPL	1297
Db	1250	ISDSDVLVFGSKTVTKNFNFKKTVVEYKKAII--BEKGLGYQ-BELINISLLCGDYT	1306
Qy	1298	TKAHLRLTSLN---LVNNDPNYKIEDLKVIKNEAGD--HOLAFSLRANNIKRLMNT-----	1347
Db	1307	IGVHGIGIVNALEIIFKAFNF--EDLKLDIVSNPFRKIDKMMYNEEQOFLNTHKNYK	1364
Qy	1348	-----PITFADYNPFFYNYNEDWRSIDKYLNKNKGNVSSHQOQAAG	1387
Db	1365	LNWIPFNNFPDREYVKCFYKPVCTDIKKFEMHVPDIKSITKFLHKTWTNISEKVL----	1420
Qy	1388	NQSGGLIQRNLKNI-----KPETFTPAL	1410
Db	1421	NVLNPLQKYNVNVRTYQSKIEDFFPLL	1448
RESULT 38			
ADJ68448			
ID	ADJ68448 standard; protein; 3225 AA.		
XX	ADJ68448;		
AC	AC		
XX	06-MAY-2004 (first entry)		
DT	Human heat mitochondrial protein as a therapeutic target SeqID254.		
DE	mitochondrial; human; screening assay; diabetes mellitus;		
XX	Huntington's disease; osteoarthritis;		
KW	Leber's hereditary optic neuropathy; LHON;		
KW	mitochondrial encephalopathy lactic acidosis and stroke; MELAS;		
KW	myoclonic epilepsy ragged red fibre syndrome; MERRF; cancer;		
KW	neuroprotective; nontropic; antidiabetic; anticonvulsant; antiarthritic;		
KW	osteopathic; ophthalmological; cytostatic.		
XX	Homo sapiens.		
OS	WO2003087768-A2.		
XX	23-OCT-2003.		
PD	04-APR-2003; 2003WO-USO10870.		
XX	12-APR-2002; 2002US-0372843P.		
PR	17-JUN-2002; 2002US-0389987P.		
PR	20-SEP-2002; 2002US-0412418P.		
XX	(MITO-) MITOKOR.		
PA	(BUCK-) BUCK INST AGE RES.		
XX	Ghosh SS, Fahy ED, Zhang B, Gibson BW, Taylor SW, Glenn GM;		
PI	Warnock DB;		
XX	WPI; 2003-845369/78.		
DR	Identifying a mitochondrial target for drug screening assays and for		
XX	treating diseases associated with altered mitochondrial function,		
PT	comprises detecting a modified polypeptide in a sample and correlating		
PT	with the disease.		
XX	Claim 1; SEQ ID NO 254; 180pp; English.		
PS	This invention relates to novel mitochondrial targets that can be used		
XX	for therapeutic intervention in treating a disease associated with		
CC	altered mitochondrial function. Specifically, it refers to a method for		
CC	identifying proteins of the human heart mitochondrial proteome that are		
CC	useful for drug screening assays, as well as therapeutic targets. The		
CC	present invention describes a method for identifying such proteins that		
CC	can be used in the treatment of various diseases associated with altered		
CC	mitochondrial function including diabetes mellitus, Huntington's disease,		
CC	osteoarthritis, Leber's hereditary optic neuropathy (LHON), mitochondrial		
CC	encephalopathy lactic acidosis and stroke (MELAS), myoclonic epilepsy		
CC	ragged red fibre syndrome (MERRF) or cancer. Accordingly, these		
CC	compositions have neuroprotective, nontropic, antidiabetic,		
CC	anticonvulsant, antiarthritic, osteopathic, ophthalmological and		
CC	cytostatic activities. This polypeptide sequence is a human heart		

CC	mitochondrial protein of the invention.		
XX	Query Match	2.8%; Score 271; DB 7; Length 3225;	
SQ	Sequence 3225 AA;	Best Local Similarity 18.8%; Pred No 3.8e-06;	
	Matches 308; Conservative 246; Mismatches 507; Indels 578; Gaps 76;		
QY	106 IPKYQVWISFEPDDEKFRGLFKLEKLEGNIAQASATKFIYLLPLDMPKALGQYS 165		
DB	1704 LSKFQSLMSEKDSLSEVQ-----DLKHQIE-GNVSKQAN-----LEATEKHNDQTN 1750		
QY	166 Y-----IVDKFNWLIHPLSNFSAQSIKPLALTRSDPTAKLNFQNDQLVW 214		
DB	1751 VTEEGTQSIPEGEODSLMSRPTCSSESVPSAKSANPAVSKD-----FSSHDEINN 1803		
QY	215 YLKEFFDLEALKANIRLQTAADPFSEKGNLVDPVYSFIRPNQNKQWASDLNODQKTVRL 274		
DB	1804 YLQQ-----IDQLKERI-----AGLEEEK-----QNKKEFSQTLNEKNTLLS 1841		
QY	275 YLRTFSPQAKTILKDYKYKDETFLSIDLKASNGTSL-----PANENDLKQDLDVLDLV 330		
DB	1842 QISTK-DGELK-MLQEVTKWNLNQIQEELSRTVKLTETAEEKXDDLEERLWNLQAL 1899		
QY	331 SDYFGQSETITS-----NSQVPV-----PASERSUKORVKFKDQOKPRIEK 374		
DB	1900 NGSIGNYCDQVTDQAQKNELLESEMKNLKCVCSELEEEKQQLVKBKTKVSEIRKEYLEK 1959		
QY	375 F-SLYEYDALSFYSOLOELV-SKPNISIKDLVN-----ATLARNLR----- 412		
DB	1960 IQQAQKEPGNKSHAKELQELKKEQOEVLQKQDCIRYQEKISALERTVKALEFVOTESQ 2019		
QY	413 -----FSLGKYNFLDLDLASHLDYFVLVSKAKIKQSSITKJFI 451		
DB	2020 KOLEITKENLAQAVEHKKQAQELASFVLLDDTQSEA-ARVLADNLKLKELQSNKESV 2078		
QY	452 ELPIK---ISLKSIIQDQPNITKTFEKEVTPLDNFR---DVEIEKAFGLLYPGVN--- 503		
DB	2079 KSMQKQKDEDLRLLEQAEEHLKE-KKNQOEKLDALREKHVLEBETTIGEIQVTLNKKD 2136		
QY	504 EELEQARK-----AQASPEK-----EKSKKGLAKERS-----QOKE 534		
DB	2137 KEVQQLQENLDSTVTQLAATKMSLSQDDRDVIDEAKKWKRFSDAIQSKEEIRLKE 2196		
QY	535 ENSKAINNQ-----EGLEEDDNITERLPENSPFIQYOQENAGLGASDPKPY 579		
DB	2197 DNGSVLKDLQRQMSIHWEELKINISLREHDKQIWESKAO-TEVOLQOKVCDTLQGENKE- 2254		
QY	580 MIKDVQNRYYLAKSQIQELIKAKDYTKLAKLLSNRHTYNIURLKEQLFDVNPRIYPSR 639		
DB	2255 LLSQLEBTRHLYHSSQ-----NELAKLESSELS-----LKDQLTDL-----SN 2292		
QY	640 DIEKAKFVLDKTEKNKYQIYSSASPVFQNKWSLFGVYRYLLGLDPKQTHIELVKLGOKA 699		
DB	2293 SLEKCK-----EKGNGLEGIIRQ-----QEA 2313		
QY	700 GLQFE--GYENLPSDFNLE-----DLKNIRIKTFLFSQKDNFKLSLLDFNNYD 746		
DB	2314 DIQNSKFSYEQLTDLQASRELTSLRHEEINMKEQKIIISLSGKEAIIQVAIAELRQQHD 2373		
QY	747 GEIKAPEFGLPLPLKELRN-----SSNGSGSNGSNSPWEQEIISOFKQONLSNQDLAQ 802		
DB	2374 KEIKE-----LENLSQEEENIVLEENKKAVDKTNQ--LMETLTKTKKENIQQAQLDS 2427		
QY	803 FSTKI-----WEKITGDENEFD-----ONNRL----- 824		
DB	2428 FVKSMSSLQNDRIIVGDYQOOLEERHLSITILEKDLQIQEAAENKLSKEIRGLRSHMDD 2487		
QY	825 -----QYK-----LLKDLQESWI-----NKTRDNLWYTYLGDKLKVRP 857		
DB	2488 LNSENAKDLAELIYREDLNQVITIKDSQKQLLEVQLQONKELENKY-AKLEEKLESE 2546		
QY	858 KNN--LEAKF-----RQISNLQF----- 873		
DB	2547 EANEELRRSFNALQBEKQDLSKEISLKVSIQSOLTRQVTALEBEEGLGLYHQAQKVKKEE 2606		
QY	874 --LLTAFYTS-----AALSNMNYIQDSGAKSTIIFEEIAELDPKVKYKVGADVYQLK-F 925		
DB	2607 VHLRSALFSSSQKRIAELEELVVCVQKEAAK-----KVEIEDKLK-----ELKHL 2653		
QY	926 HYAIFDDNAGKFNQEVIRSSRTIYLTGSGSKULEADTIDQLNQAVKNAPLGLQSF--- 982		
DB	2654 HHDAGIMRNETETABERVAELARDLV-----EMEQKLL-MVTKENKGLTAIQISFGRS 2705		
QY	983 -----YLTERFGVFOKLATSIAVQHKQKKTLPKLLANDGYTILHDKLKKVIPQ 1033		
DB	2706 MSSLQNSRDHANEELDELKKYDASLKAELAKEQGL---LARE-----RDALLSETAFS 2757		
QY	1034 ISSSEKQWFE-GKLNQNGSQNVNVSTFGSIIESPY-----FS---TNFQEDAD----- 1079		
DB	2758 MNSTENSLSHLEKLNQQLLSKDEQLLLSSQLEDSYNQVQSFASKAMASLQNERDHLWNE 2817		
QY	1080 -----LDQDQDSDS-----ROGNNSLDNQBAGLLKQKLAILLGNQFYQYQ 1120		
DB	2818 LEKFRKSEBGKQSAAPSTSPAQVQSLKKAMSSLQNDRLRLLKE-----LKNLQOQYQLQ 2872		
QY	1121 QN-----DKIEIEFINVE-----KVSLSF 1141		
DB	2873 INQETELHPLKAQLOEYQDQTKAFQIMQBELRQENLSWQHEDQLRMKSSWEIHERM 2932		
QY	1142 RVEFKLAKTLENG-----KTIRVLSDETMSLIVNTTIKTPMSAVPE-----VFD 1188		
DB	2933 KEQYLMASIDKQOQLSHLQNLRELSSSSQTPKQVQROQASPTSPDGSQNLVYE 2992		
QY	1189 TKWFOYDRTPLAAKTKFVLKPK---DQIPVDGSGNISDKWLASIPLVIHQOMLRLSPV 1245		
DB	2993 TELL-----RTQLNDSLKEIHQELRIQOLNSNFSQLLEKNTLSIQLCDTSQSLR--- 3043		
QY	1246 VKTIRELGKLTQQQ-----QOQOQOQOQP-----QKXAV-RKEE 1281		
DB	3044 -----ENQOYHGDLLNHCAVLEKQVQELQAGPLNIDVAPGAPQEKNGVHRKSDP 3092		
QY	1282 LETYNPKOEFNLNPLTKAHLRTLNLVNDPNYKIEDLKVKNEAGDHQL----AFSLR 1337		
DB	3093 EELRFPQQSFSQAQ-----QQLCNT--RQEVNELRKLLEERDQORVAENALSVA 3140		
QY	1338 ANNIKRL-----MNTPI 1349		
DB	3141 EQQIRRLSEHSEWSSRTPI 3159		
RESULT 39			
ABU42656	ABU42656 standard; protein; 6641 AA.		
ID	XX		
AC	ABU42656;		
XX	DT 19-JUN-2003 (first entry)		
DE	Protein encoded by Prokaryotic essential gene #28183.		
KW	Antisense; prokaryotic essential gene; cell proliferation; drug design.		
XX	Staphylococcus epidermidis.		
XX	WO200277183-A2.		
PN	03-OCT-2002.		
PD	21-MAR-2002; 2002WO-US009107.		
PF	21-MAR-2001; 2001US-00815242.		
PR	06-SEP-2001; 2001US-00948993.		
PR	25-OCT-2001; 2001US-0342923P.		
PR	08-FEB-2002; 2002US-00072851.		
PR	06-MAR-2002; 2002US-0362699P.		

XX PA (ELIT-) ELITRA PHARM INC.

XX PI Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;

XX PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;

XX DR WPI; 2003-029926/02.

XX DR N-PSDB; ACA46526.

XX PT New antisense nucleic acids, useful for identifying proteins or screening

XX PT for homologous nucleic acids required for cellular proliferation to

XX PT isolate candidate molecules for rational drug discovery programs.

XX PS Claim 25; SEQ ID NO 70580; 1766pp; English.

XX CC The invention relates to an isolated nucleic acid comprising any one of

XX CC the 6213 antisense sequences given in the specification where expression

XX CC of the nucleic acid inhibits proliferation of a cell. Also included are:

XX CC (1) a vector comprising a promoter operably linked to the nucleic acid

XX CC encoding a polypeptide whose expression is inhibited by the antisense

XX CC nucleic acid; (2) a host cell containing the vector; (3) an isolated

XX CC polypeptide or its fragment whose expression is inhibited by the

XX CC antisense nucleic acid; (4) an antibody capable of specifically binding

XX CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular

XX CC proliferation or the activity of a gene in an operon required for

XX CC proliferation; (7) identifying a compound that influences the activity of

XX CC the gene product or that has an activity against a biological pathway

XX CC required for proliferation, or that inhibits cellular proliferation; (8)

XX CC identifying a gene required for cellular proliferation or the biological

XX CC pathway in which a proliferation-required gene or its gene product lies

XX CC or a gene on which the test compound that inhibits proliferation of an

XX CC organism acts; (9) manufacturing an antibiotic; (10) profiling a

XX CC compound's activity; (11) a culture comprising strains in which the gene

XX CC product is overexpressed or underexpressed; (12) determining the extent

XX CC to which each of the strains is present in a culture or collection of

XX CC strains; or (13) identifying the target of a compound that inhibits the

XX CC proliferation of an organism. The antisense nucleic acids are useful for

XX CC identifying proteins or screening for homologous nucleic acids required

XX CC for cellular proliferation to isolate candidate molecules for rational

XX CC drug discovery programs, or for screening homologous nucleic acids

XX CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,

XX CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is encoded by one of

XX CC the target prokaryotic essential genes. Note: The sequence data for this

XX CC patent did not form part of the printed specification, but was obtained

XX CC in electronic format directly from WIPO at

XX CC ftp.wipo.int/pub/published_pct_sequences

XX SQ Sequence 6641 AA;

Query Match 2.8%; Score 270; DB 6; Length 6641;

Best Local Similarity 18.7%; Pred. No. 1.3e-05;

Matches 411; Conservative 310; Mismatches 772; Indels 708; Gaps 103;

QY 61 NFTSDYQSVKALLNGKTFDPKSSSEFTDFVSKFDFLTNNRTVLEIPKQYVVISFSP 120

DB 2741 NTSNSRDDIKQIEQAKALNDAMKLEQVQKQGVHAN-----SDYTN 2785

QY 121 DDKERPLGFLHKE-----KLEDGNI-AQSATKFTYLLPLDMPKAAL-----GOYSYIVD 169

DB 2786 DSAQKDAYNALKQAEIDIINSSNPMLNAQDITN-----ALNNIQAQDNLHGAQKLQD 2840

QY 170 KNFNLIILHPLNFS-----AQSILKPL-----ALTRSSDFIAKLN 204

DB 2841 KNTTNAQGNLNLHNPQKDALIQAINGATSRSDQVAEKLKEABALDEAMKQLEDQVND 2900

QY 205 QFNW-----QDELWVYLEKFFDLEALKANIRLQATDFSPKGNLVDPPFVYSFTRNQ 256

DB 2901 QISNSSPFINESDKQTYNDK---IQAAKEIN-QTSNPTLDKQKIADTL-----Q 2948

QY 257 NQEWASDLNQDQKTVRLY-----LRTFSPQAKTILKDYKYKDETFILSSIDL 304

DB 2949 NIKDAVNNLHGDQKLAQSKQDANNQNLHLLDLEEQKNHFKPLINNADTRDEVNQLEIFA 3008

QY 305 KASNG-----TSLFAN-ENLDKQDQDVLDDLVSYFQGSQTTSNSQ 346

DB 3009 KQLNGDMSTLHKVINDKQIQHLSNINADNDKKQNYDNAKEADLIHNPDTL----- 3063

QY 347 VKPVPASERSLKDRVKFKQDQKPKRIEKSLEY----DALSFYSOQLQEL-VSPKNSIKOL 402

DB 3064 -----DHKALQDLLN-KIDQAHNELNGESRFKQALDNALNDIDSLNSLNVPOQTVDKN 3116

QY 403 VN-ATLARNLRFSLGK-----YNFLFDDLASHLDYFVLVSK 437

DB 3117 INHVTTLESIAQELQAKELNDAMKAMRDSIMNQEQIRKNSNTYNEDLAQONAYHAVDK 3176

QY 438 AK-----IKOSSITKFLFELPIKISLKSSILGDQE-PNITKTUFEKEVTFKLDNFRD 488

DB 3177 INNIIGEDNATMDPQIIKQATQDINTAI---NGLKGDQKLQDAKTDAKQDIT-----NFT- 3228

QY 489 VEIEKAFGLLYPG-----VNEELBOARKAQRASPEK-----EKSKGLKAFSQQKEE 535

DB 3229 -----GLTEPOKQALENIINOQTSRANVAKQLSHAKFLNGKMEELKAVAKASLYR-Q 3280

QY 536 NSKAINNQGLEEDDNIATERLPENSPIOYQOENAGLGASDPKPYMIKDV-----QN----- 586

DB 3281 NSNYIN-EDVSKEAVEQAIAGQOEIINSENNPTI-SSTDINRTIOEINDAEQNLHGDN 3337

QY 587 ---QRYVLAKSQIQEL--IKADYTKL-----AKLLSNRHTYNTISLRKEQ 627

DB 3338 KLRQAEIAKNEIQNLGLNSAQITKLIQDIGRTTTPKPAVTQKLEAKAINQAMQQLKQS 3397

QY 628 LFDVNPRIIPSRDIEKAKFVLDKTERKQYQIYSSASPVQNKWSLFGYRYLLGLBPQK 687

DB 3398 IADKDATLNSSNVINE-----DSEKKLAYDNVSAEQELI-NQLN-----DPTM 3440

QY 688 TIHELVKLGOK---AGLQFSGYENL---PSDENLEDLKNIRIKTFLPSQKDNF-----KLS 737

DB 3441 DISNIQAITQKVIOAKOSLHGANKLAQNAQDSNL-----IINQSTNLNDKQKA 3489

QY 738 LLDFFNNYDGEIKAFEFGLFLFLPKELRRNSNGG-----SONSNSFWOEIISQPKD- 791

DB 3490 LNDLINHAQTKQVAE-----IIAQANKLNENGTTLKTLVEEQSNVHQSKYINEDPQV 3543

QY 792 QNLSNQDLAQFSTKIWEKIIGDENFDQNRLL-----QYLLKQ-----LQESWINK 839

DB 3544 QNIYND-----IQKGREILNGTTDDVLNNNKIADAIQNTIHLTKNDLHGDQKLQAKQDA 3598

QY 840 TRDNLNLYTYLGDKLKVPKKNLEA--KPROISN-----LQELLTAPYTSAAALSNN--- 887

DB 3599 TNELNLYTLNLSNQSEHDEINSPSRTEVSNDLNHAKALNEAMRQLENEVALENSVYK 3658

QY 888 -----WNYQDSGAK-----STI-----IFEEIABLDPKVKEKVGADYVQ 922

DB 3659 LSDFINEDEAAQNEYSNALQAKADIIINGVPESTLDKATIEDALLELQNARESLHGEQKLQ 3718

QY 923 LKFHYAIGFDNAGKFNQEVIRSSRTIYLYKTSKSKLE--ADTIDQLNAQV----- 973

DB 3719 EAKQOAVAEIDNLQALNPGQV-LAEKTLVNQASTKPEVQEAQLOKAKELNEAMKALKTEIN 3777

QY 974 -----NAPLGLQSEFYLDTERFVGFKLATSQVQKQKELTPKPLKNDGYTL 1021

DB 3778 KKEQTKADSRVYNADSGLOQANYNSALNYG-SQIAT-----TOPPELNKDVINR 3825

QY 1022 IHDKLKKEPVIQIISSSPEKDFEGKLNQNGSQNVNVSTFGS-IIESPYFTNFQEDADL 1080

DB 3826 ATQITK-----TAENNLNGQSLAEAKSDGNQSTIEHLQGLTQSQDKQH 3869

QY 1081 DODGQDSDRQ-----NNS--LDNQEAGLLKQKLAILLGNQFIQYQQNDKEIB--FEII 1131

DB 3870 DLINQATQKQVDDIDVNNSKOLDNSM-----NQ-LOQIVNNDNTVKQNSDFI 3915

QY 1132 NVEKVSLSFRVEFKLAKTLEDNGKTRIVLSDETMSLVNTTIEKTPMSAVPEVFTDKW 1191

DB 3916 NEDSQQDAYNHAIQAAKL-----ITAHPTIMDKNQ 3947

QY 1192 V5QYDPTPLAAKTKFVLKFKDQI-PVDGSGNIS-DKWLASIPLVIHQMLRLSLPVVK-- 1247

This Page Blank (uspio)